

# Wissenschaftliches Publizieren im digitalen Zeitalter



Herausgeber:  
Prof. Dr. Peter Schirnbacher  
Direktor Computer- und Medienservice  
Telefon: (030) 2093-70010  
schirnbacher@cms.hu-berlin.de

Redaktion:  
Prof. Dr. Peter Schirnbacher  
Telefon: (030) 2093-70010  
schirnbacher@cms.hu-berlin.de

Layout und Satz:  
Katrin Cortez de Lobão  
Telefon: (030) 2093-70032  
cortezka@cms.hu-berlin.de

Druck:  
Druckhaus Berlin-Mitte GmbH

Redaktionsschluss:  
1.4.2009

Postanschrift:  
Humboldt-Universität zu Berlin  
Computer- und Medienservice  
Unter den Linden 6  
10099 Berlin

Sitz:  
Erwin Schrödinger-Zentrum  
Rudower Chaussee 26  
12489 Berlin

Telefon: (030) 2093-70010  
Telefax: (030) 2093-2959  
edoc.hu-berlin.de/cms-journal/

Titelmotiv: Gutenbergportrait  
des Künstlers Dietmar Gross,  
Gutenberg-Museum Mainz  
Verfremdung des Portraits:  
Katrin Cortez de Lobão

Die Zukunft des Wissenschaftlichen Publizierens Michael Seadle	3
Einführung in das Heft Peter Schirnbacher	5
Das wissenschaftliche Publizieren – Stand und Perspektiven Peter Schirnbacher, Uwe Müller	7
Geschichtlicher Abriss – Eckdaten der Entwicklung des Publizierens	13
Möglichkeiten und gegenwärtige Grenzen des elektronischen Publizierens Peter Schirnbacher	14
Publizieren im Open-Access-Modell Stefan Gradmann	20
Überblick zum goldenen Weg zu Open Access in Deutschland und international Karin Weishaupt	24
Auf dem goldenen Weg zum edoc-Server – Open-Access-Zeitschriften, Schriftenreihen und Sammelbände an der Humboldt-Universität Uwe Müller	29
Peer Review bei Open-Access-Zeitschriften Uwe Müller	33
Die neue Q-Serie: eine Ergänzung des edoc-Angebotes Susanne Dobratz, Karin Schröder	38
Der „grüne Weg zu Open Access“ – Aktuelle Entwicklungen in Deutschland Uwe Müller, Peter Schirnbacher	41
Wie entsteht ein Institutional Repository? – Eine systematische Hinführung in acht Schritten Susanne Dobratz, Uwe Müller	47
Sieben auf einen Streich – Argumente für eine Open-Access-Publikation auf dem edoc-Server der HU	55

eSciDoc – Das Repository-Konzept der Max Planck Digital Library Malte Dreyer, Ulla Tschida	56
Rechtliche Probleme bei Postpublikationen und die Ergebnisse des DFG-Projekts „Open Access Policies“ Dennis Zielke	60
Signatur und Zeitstempel zur Wahrung von Authentizität und Integrität Niels Fromm	63
Backup-Strategie für den Dokumentenserver Niels Fromm	67
Open Access – Verfügbar ist noch nicht präsent Robin Malitz	70
Ein Zitationsindex elektronischer Dokumente in institutionellen Repositorien Frank Havemann	74
Impact von Open-Access-Publikationen Sabine Henneberger	77
Creative Commons Manuel Klatt-Kafemann	81
Der edoc-Hostingservice für andere Dokumentenserver Susanne Dobratz	86
EoD bringt alte Bücher nach Hause Patricia Fechner	90
Grundfragen der digitalen Langzeitarchivierung für den edoc-Server Susanne Dobratz	93
Service für Autorinnen und Autoren Maxi Kindling, Peter Schirmbacher	99
Elektronische Publikation von Dissertationen und Habilitationsschriften an der Humboldt-Universität zu Berlin Sabine Henneberger, Karin Schröder	102
Hinweise zu Open-Access-Veröffentlichungen auf dem edoc-Server der Humboldt-Universität zu Berlin	105
Eine gute Wahl: Wissenschaftliche Arbeiten mit LaTeX verfassen Joanna Ryćko, Manuel Klatt-Kafemann	106
Literaturverwaltung zur Unterstützung des wissenschaftlichen Publizierens Maxi Kindling, Matti Stöhr	109
Weiterbildung Personalia	104 III

# Editorial

## Die Zukunft des Wissenschaftlichen Publizierens

Die Veränderungen in der Welt des wissenschaftlichen Publizierens lassen sich kurz auf zwei Punkte reduzieren: von Papier zu digitalen Ausgaben und vom kostenpflichtigen Abonnieren zum Open Access. Das heißt aber nicht, dass man ohne weiteres sagen kann, dass digitale Open-Access-Veröffentlichungen die Zukunft sind.

In den Naturwissenschaften ist die Mehrzahl der Wissenschaftler der Auffassung, dass wissenschaftliche Werke in digitalen Formaten erscheinen sollten. In einem Fachbereich wie Physik sind Pre-Print Archive wie ArXiv ohne Frage wichtige Ressourcen. Für Wirtschaftswissenschaftler in vielen Ländern ist das Social Science Research Network ebenso ein Muss, wenn man in einer gut angesehenen wissenschaftlichen Zeitschrift veröffentlichen möchte. Was diese und ähnliche Fachbereiche verbindet, ist, dass sie zeitschriftenorientiert sind. Die buchorientierten Geisteswissenschaftler stimmen dieser Herangehensweise in der Regel nicht zu. Sie erzeugen ihre Werke natürlich mit Textverarbeitungssystemen, aber sie lieben die Festigkeit des gedruckten Endprodukts. Gegen die Liebe kann man nur schwer argumentieren. Ob ihresgleichen sich vermehrt, annähernd gleich bleibt oder endlich ausstirbt kann ich nicht vorhersagen, aber ich würde auf Aussterben wetten. Die neue, noch fehlerhafte Generation von eBook-Readern vereint fast alle Vorteile von Papierbüchern, nur gegen die Gewohnheit kommen sie noch nicht an.

Papierbasierte gedruckte Werke sind textorientierte Werke, heutzutage oft mit Abbildungen versehen. Solch eine Beschränkung ist für digitale Werke nicht nötig. Eine elektronische Publikation kann Ton, Video und interaktive Spiele einschließen. Die Popularität dieser Formate liegt jedoch überwiegend im Unterhaltungsbereich und viele Menschen lehnen diese Formate gerade deshalb für wissenschaftliche Zwecke einfach ab.

Immerhin steigen jedoch die Benutzerzahlen. Das „Last Lecture of Randy Pausch“ ist als Symbol der Änderung interpretierbar. Randy Pausch, Professor an der Carnegie Mellon University, interessiert sich für Videos und Spiele und für den Unterhaltungsbereich. Er studiert es mit wissenschaftlichem Ernst. Seine „Last Lecture“ hat beides: emotionale Wirkung und Forschungsinhalte. Es zeigt, wie diese neuen Medien auch Ideen aus einer Universität effektiv kommunizieren könnten.



Abb. 1: Digitale Texte und die Zukunft: Die Berliner Firma txtr kündigt den txtr reader für Herbst 2009 an.

Es gibt bis jetzt, so weit wie ich weiß, keine Zeitschrift mit einem Peer Review für multimediale Werke, aber solch eine Zeitschrift könnte man unter technischen Gesichtspunkten problemlos gründen. Eine Multimedia-Datei ist als digitales Objekt nicht anders als die PDF-Dateien, die die meisten Online-Zeitschriften benutzen. Man braucht geeignete Software, um ein PDF lesen zu können und man braucht geeignete Software, um Videoformate sehen und verstehen zu können. Heute ist auf den meisten Rechnern die geeignete Software schon installiert. Woran es bei wissenschaftlichen Publikationen in multimedialen Formaten mangelt, ist allein die Akzeptanz. In den

USA bekommt man mit Video-Veröffentlichungen an den besten Universitäten „tenure“ nur in wenigen Fachgebieten (z. B. Fine Arts). In Deutschland bekommt man auf diese Weise sicher keine Professorenstelle. Die heutige Generation der Universitätsleitungen ist in einer textlastigen Welt aufgewachsen und traut sich noch nicht vorzustellen, wie Wissenschaftler ihre Ergebnisse besser in multimedialer Form als in Text darstellen könnten. Die heutige Generation der Universitätsforscher ist ebenso mit Textsystemen statt Video-Editoren aufgewachsen. Sie denken überwiegend in Worten, nicht in Bildern und Ton und Bewegung. Es ändert sich vielleicht, wenn Video-Editoren ein Standardteil von Open Office sind.

Fast niemand zweifelt, dass immer mehr wissenschaftliche Werke in digitalen Formaten erscheinen. Das Gleiche kann man mit Sicherheit nicht für Open Access sagen. Viele Verlage haben sich für digitale Formate entschieden, aber sie erschießen sich nicht wegen einer mächtigen Open-Access-Bewegung. Hier ist der Kampf um die Zukunft noch zu entscheiden.

Open Access soll im Prinzip leicht definierbar und erkennbar sein: alles, was kostenlos zugänglich ist. Es gibt jedoch Werke und Verlage, die in einem grauen Mittelfeld liegen. Einige Verlage bieten zum Beispiel den kostenlosen Zugang zu einigen Werken an, manchmal auch nur zeitlich begrenzt. Man nennt solche Werke „loss leaders“. Ziel ist es, Leser an dem Werk so zu interessieren, dass sie sich letztlich den Zugang zum ganzen Werk kaufen. Kostenlos, ja. Open Access im eigentlichen Sinne, nein.

Die Romeo/Sherpa-Auflistung von Verlagen zeigt, wie Open-Access-freundlich sie sind. Die Freundlichkeit versteht sich in dem Sinne, dass es Autoren gestattet ist, ihre Werke in Open-Access-Repositorien hochladen zu können. Der Verlag, mit dem ich als Editor eng zu-

sammenarbeite, Emerald, hat eine dem Open-Access-Gedanken aufgeschlossene Politik: Man darf ohne weitere Genehmigungen sein eigenes Werk in einem Repository speichern, mit der einzigen Bedingung, dass man angibt, dass es ursprünglich in einer Emerald-Zeitschrift veröffentlicht wurde. Der Präsident ist kein überzeugter Open-Access-Freund, sondern ein fähiger Businessman. Er weiß, wie wenige Autoren dieses Angebot annehmen und er weiß auch, dass der Verlag mit dieser Politik auch solche Autoren gewinnt, die aus Prinzip nur in Open Access oder Open-Access-freundlichen Zeitschriften veröffentlichen. Selbst wenn deutlich mehr Autoren ihre Werke in Open Access Repositorien einstellen, bleibt unklar, ob der Lesende diese Werke wirklich findet. Es ist leider der Fall, dass viele Repositorien nicht besonders benutzerfreundlich sind und dass die meisten schlecht mit einander vernetzt und indiziert sind. Auch die Benutzer sind erst zu überzeugen, dass sie tatsächlich auch in Open-Access-Quellen suchen.

Ob Green Road oder Golden Road das beste Open-Access-Businessmodell ist, wird breit diskutiert. Der Teufel steckt diesmal nicht im Detail, sondern in den Kosten. Es muss jemanden geben, der die Kosten der Veröffentlichung übernimmt. Beim Green Road bezahlen die Einrichtungen, die hinter den Repositorien stehen. Es sind normalerweise die Universitäten. Begründet wird diese Haltung mit dem Argument, dass durch die Repositorien die Kosten gespart werden, die auszugeben wären, wenn man die in den Universitäten entstandenen wissenschaftlichen Arbeiten von den Verlagen zurückkaufen müsste. Die Fragen sind: Wer organisiert die Qualitätssicherung und wie werden die Repositorien benutzerfreundlicher? Es sind lösbare Probleme, deren Lösung jedoch noch aussteht. Der Golden Road hat den Vorteil, dass er ein verlagsähnliches Modell für die Qualitätssicherung benutzen kann. Er hat den Nachteil, dass kein Geisteswissenschaftler die ausreichenden Mittel hat, für seine eigenen Veröffentlichungen auch zu bezahlen. Solch „vanity publishing“ hat auch, zumindest bei vielen Wissenschaftlern in

den USA, einen schlechten Ruf. Ob das Modell wirklich rentabel ist, wird noch herauszufinden sein.

Wo liegt die Zukunft des wissenschaftlichen Publizierens? Ich gehe kein großes Risiko ein, wenn ich für Texte „digital“ sage. Bei wissenschaftlichen multimedialen Publikationen wird man auf eine offeneren Denkart warten müssen, die sicher auch von einigen Pensionierungen abhängig sein wird. Ob kommerzielle Verlage oder eine der Formen von Open Access in meiner Lebenszeit die Oberhand gewinnen, hängt davon ab, ob die Open-Access-Fürsprecher ein Businessmodell finden, das wirklich konkurrenzfähig ist. Ich wäre dafür, aber ich zweifle, ob die kommerziellen Verlage schon das Gefühl haben, dass sie in Gefahr seien.

Prof. Michael Seadle PhD  
Geschäftsführender Direktor des  
Instituts für Bibliotheks- und  
Informationswissenschaft  
Professur für Digitale Bibliotheken

# Einführung in das Heft

Peter Schirmbacher | schirmbacher@cms.hu-berlin.de

Im Herbst 1991 haben wir innerhalb des damaligen Rechenzentrums der Humboldt-Universität längere Zeit um den Charakter einer Schriftenreihe gerungen, die dann erstmalig am 06.11.1991 unter dem Titel „RZ-Mitteilungen“ erschien und heute mit der 32. Ausgabe des cms-journals (die Namensänderung erfolgte als Reaktion auf die Namensänderung des Rechenzentrums in Computer- und Medienservice) seine Fortsetzung findet. Es war damals die Zeit, als immer mehr Mitgliedern der Universität die Möglichkeit eröffnet wurde, direkt auf das universitätsinterne und internationale Rechnernetz zuzugreifen. Die dort abgreifbaren, aber vor allem durch uns einstellbaren Informationen an unsere Nutzer waren deutlich aktueller als jedes Mitteilungsblatt, was wir hätten erarbeiten, drucken und verteilen können. Die aus den Jahren davor geprägte Kommunikation zwischen dem zentralen Rechenzentrum und seinen vielen Nutzergruppen war überholt und ist, wie wir heute bestens wissen und praktizieren, durch das WWW abgelöst worden. Trotzdem haben wir diese Schriftenreihe ins Leben gerufen, um damit über die aktuelle und teilweise schnelllebige Web-Nachricht hinaus einzelne Themen detaillierter für die Nutzerschaft aufzuarbeiten.

Die Absicht ist die Bereitstellung einer ergänzenden Lektüre, die sich nun in dem vorliegenden Heft mit dem „Wissenschaftlichen Publizieren im digitalen Zeitalter“ befasst. Aus unserer Sicht ist das Thema hoch aktuell und berührt nahezu jede Mitarbeiterin und jeden Mitarbeiter der Universität. Das wissenschaftliche Publizieren findet, so könnte man

im übertragenden Sinne sagen, am Computer statt. Jeder Wissenschaftlerarbeitsplatz ist mit einem Computer ausgestattet und das „geistige Werk“ findet dort seinen „ersten“ Speicherplatz, wird bearbeitet, möglicherweise über das Netz vorab verschickt, um es mit Fachkolleginnen und Fachkollegen zu diskutieren und kommt erst dann zur eigentlichen Veröffentlichung, die in der Mehrzahl der Fälle leider und ausschließlich in Papierform erscheint.

Das vorliegende Heft ist eine weitere Zusammenfassung (siehe auch Heft 27 des cms-Journals von 2005) der bisherigen Arbeiten der gemeinsamen Gruppe des Computer- und Medienservice und der Universitätsbibliothek der Humboldt-Universität zu Berlin zum „Elektronischen Publizieren“. Die Gruppe existiert, seit im Jahre 1997 erstmalig an der Humboldt-Universität der Versuch unternommen wurde, der Veröffentlichungspflicht bei Dissertationen durch eine elektronische Version zu entsprechen. Dem Anliegen, wissenschaftliche Veröffentlichungen auch elektronisch in ihrem Volltext vorzuhalten, konnte damals wie heute nur entsprochen werden, indem dieser Service für die Universität als Gemeinschaftsaufgabe der einschlägigen Infrastruktureinrichtungen verstanden wird. Es ging und geht um die Bündelung des Know-hows einer wissenschaftlichen Bibliothek mit dem eines wissenschaftlichen Rechenzentrums. Seit dieser Zeit ist rund um den edoc-Server (siehe <http://edoc.hu-berlin.de>) ein umfassender Service aufgebaut worden, dessen Ziel die möglichst umfassende Unterstützung der einzelnen Phasen des Publikationsprozesses

darstellt. Die Inhalte des Servers, gegenwärtig immerhin fast 8.000 Dokumente, sind breit gefächert. In ihrer Masse werden sie durch Hochschulschriften wie Dissertationen und Habilitationen bestimmt. Es werden jedoch auch acht wissenschaftliche Zeitschriften gehostet, Konferenzproceedings angeboten und für immer mehr Institute der Universität ihre jeweiligen Schriftenreihen und Sammelbände veröffentlicht. Ein besonderes Ziel der Arbeitsgruppe ist es, in gleicher Weise den Prozess des Publizierens an der Universität zu unterstützen, den Autorinnen und Autoren technische Hilfestellungen zu geben und dabei insbesondere dem Open-Access-Gedanken noch deutlicher zum Durchbruch zu verhelfen.

Die Facetten der Arbeiten der Gruppe sind äußerst vielschichtig und spiegeln sich entsprechend in den Beiträgen zu diesem cms-journal wider. Sie reichen von der theoretisch geprägten Auseinandersetzung mit den Grundlagen des Publizierens, wobei das elektronische Publizieren zweifelsfrei im Mittelpunkt steht, über die Beschäftigung mit den Anforderungen an eine elektronisch verfügbare wissenschaftliche Zeitschrift und dem damit verbundenen „Forderungskatalog“ eines Herausgebers, der bestmöglich technisch unterstützt werden möchte, bis zur detaillierten Bestimmung der notwendigen Servicequalität eines wissenschaftlichen Repositories, wie es der edoc-Server darstellt. Das Heft gliedert sich in einen theoretischen Teil, in dem der Stand und mögliche Perspektiven des elektronischen Publizierens aufgearbeitet werden. Es wendet sich anschließend den Grundzügen des Publizierens nach den Prinzipien des Open Access zu, um dabei den beiden hauptsächlichen Ausprägungen, dem „Golden Road“ bzw. dem „Green Road“ des Open Access mit jeweils mehreren Beiträgen zu folgen. Zwangsläufig liegt bei vielen Beiträgen der Schwerpunkt auf der Vermittlung der vielfältigen Erfahrungen, die im Verlauf der Jahre durch den Betrieb des Repositories und die Bearbeitung einer großen Zahl von Projekten der angewandten Forschung, finanziert durch die Deutsche Forschungsgemeinschaft (DFG), das Bundesministerium

für Bildung und Forschung und die Europäische Gemeinschaft, gesammelt wurden.

Mit diesem Heft wenden wir uns an die Mitarbeiterinnen und Mitarbeiter und damit Autorinnen und Autoren der Humboldt-Universität, indem wir hoffen, ihnen eine Handreichung als Unterstützung im wissenschaftlichen Publikationsprozess zu geben, das ein oder andere Vorurteil gegenüber dem elektronischen Publizieren auszuräumen und sie natürlich auch für die Nutzung des edoc-Servers etwas zu begeistern. Wir möchten die Wissenschaftlerinnen und Wissenschaftler dieser Universität ansprechen, die gegenwärtig als Herausgeber wissenschaftlicher Zeitschriften fungieren bzw. beabsichtigen, in der nächsten Zeit eine Zeitschrift ins Leben zu rufen, um sie auf die Möglichkeiten des edoc-Servers auf diesem Gebiet hinzuweisen. Natürlich ist das Heft auch gedacht für unsere Kolleginnen und Kollegen, die in anderen Einrichtungen mit uns vergleichbare Aufgaben zu erfüllen haben und last but not least soll das Heft als eine Unterstützung bei der Ausbildung von Studierenden im Modul „Elektronisches Publizieren“ des Instituts für Bibliotheks- und Informationswissenschaft dienen.

uns als Folge davon über mangelnde Aufgaben nicht beklagen können. Umso mehr gilt mein Dank allen Kolleginnen und Kollegen, die hier als Autorinnen und Autoren in Aktion getreten sind, für ihre Mühen und ihr Engagement.

Besonders freue ich mich jedoch, dass wir auch Kolleginnen und Kollegen außerhalb der Arbeitsgruppe „Elektronisches Publizieren“ gewinnen konnten, sich mit Beiträgen an diesem Heft zu beteiligen und es dadurch zu bereichern. Zugegebenermaßen teile ich nicht jede der geäußerten Meinungen, aber gerade darin sollte ja der Reiz einer wissenschaftlichen Veröffentlichung liegen.

Den Leserinnen und Lesern des cms-Journals wünsche ich eine angenehme Lektüre und äußere die Bitte, mit Anregungen oder kritiken nicht zu sparen, sondern uns Ihre Meinungen mitzuteilen.



Abb. 1: Arbeitsgruppe Elektronisches Publizieren

Die Resonanz auf das Wirken der Arbeitsgruppe in der Universität und darüber hinaus ist sehr gut, sodass wir

# Das wissenschaftliche Publizieren – Stand und Perspektiven

Peter Schirmbacher | schirmbacher@cms.hu-berlin.de

Uwe Müller | u.mueller@cms.hu-berlin.de

## Die wissenschaftliche Publikation

Die Entstehung der Schrift, die Erfindung des Papiers und die Nutzung der „beweglichen Letter“ zum Buchdruck stellten unbestritten Grundlagen und Meilensteine sowohl für die Kommunikation der Menschheit insgesamt als auch für den wissenschaftlichen Gedankenaustausch im Besonderen dar. Die 1665 in Paris und London gegründeten weltweit ersten wissenschaftlichen Fachzeitschriften sind in ähnlicher Weise historische Marken für die Wissenschaft. Durch die relativ jungen Möglichkeiten digitaler Verarbeitung und Verbreitung wissenschaftlicher Ergebnisse haben wir es mit einer ähnlich einschneidenden Entwicklung zu tun. Sie hat die wissenschaftliche Kommunikation zunächst fast unbemerkt erreicht, verändert sie jedoch nun bereits seit mehr als einem Jahrzehnt mehr oder weniger systematisch und wird sie in der Zukunft zweifelsfrei erheblich prägen. Der Beitrag beschreibt zunächst das Phänomen der wissenschaftlichen Publikation bzw. des wissenschaftlichen Publikationsprozesses, um anschließend potentielle Entwicklungen in ihren Ansätzen aufzuzeigen.

## Funktionen des wissenschaftlichen Publizierens

Als *Publikation* wird im Allgemeinen ein öffentlich verfügbares Werk verstanden, das in einer spezifischen Repräsentation vorliegt und an ein Träger- oder Übertragungsmedium gebunden ist. Hierbei steht jedoch nicht die Verbreitung von Informationen an sich im Vordergrund,

wie dies beispielsweise auch beim Versand eines Briefes geschieht. Stattdessen ist eine Publikation „für die Öffentlichkeit, für ein mehr oder weniger anonymes Publikum bestimmt“ [1]. Als *Werke* im engeren Sinne werden dabei Ergebnisse einer geistigen Schöpfung betrachtet, die durch das Mitwirken eines oder mehrerer Urheber entstanden sind. Dazu zählen z. B. Sprachwerke, musikalische Kompositionen, Bilder, Filme und dergleichen mehr<sup>1</sup>. Nicht betrachtet werden in diesem Zusammenhang dagegen beispielsweise Ergebnisse rein mechanistischen Handelns oder der maschinell generierte Output eines Computerprogramms. Als *Publizieren* bezeichnet man den Gesamtprozess, der sich dazu geeignet, ein Werk zu einer Publikation, es also – in der Regel dauerhaft – öffentlich zugänglich zu machen.

Der wissenschaftliche Publikationsprozess wird als wesentlicher Bestandteil der Wissenschaft selbst gesehen (vgl. [2]). Er bildet nicht nur die Grundlage für Verbreitung und dauerhafte Sicherung einmal gewonnener Erkenntnisse. Er ist gleichzeitig Voraussetzung dafür, dass wissenschaftliche Forschung trotz der zeitlichen und räumlichen Verteilung der daran beteiligten Akteure aufeinander aufbauen und sich aufeinander beziehen kann. Das wissenschaftliche Publizieren ist mit der Wissenschaft an sich sowie der Produktion und dem Erwerb von Wissen eng verknüpft: Die Veröffentlichung einer wissenschaftlichen Arbeit fußt auf deren inhaltlicher Erarbeitung im Rahmen wissenschaftlicher Forschung. Andererseits kann

*Der wissenschaftliche Publikationsprozess befindet sich in einem gravierenden Wandel. Er wird zunehmend geprägt durch die Möglichkeiten der digitalen Verarbeitung und der Verbreitung wissenschaftlicher Erkenntnisse über das Internet. Damit gehen Veränderungen im Publikationsverhalten der Autorinnen und Autoren einher. Letztlich ist es das Ziel des gesamten CMS-Journals, dieses Phänomen zu beschreiben. In diesem Beitrag werden zunächst die wissenschaftliche Publikation, ihre Funktionen und Anforderungen beschrieben, bevor anschließend auf sich vollziehende Veränderungen eingegangen wird.*

*Wissenschaftliche Publikation, Publikationsprozess, Funktionen einer Publikation, Anforderungen an eine Publikation, Urheberschaft, Autoren, Verlage*

<sup>1</sup> siehe u. a. Urheberrechtsgesetz (UrhG), § 2



sie aufseiten der Rezipienten als Grundlage weiterer Wissensproduktion dienen (vgl. [1]).

Beim wissenschaftlichen Publizieren spricht man gemeinhin von einem gesellschaftlich bedingten Kreislauf, beginnend mit der Darstellung des geistigen Werkes durch die Autoren, der eigentlichen Publikation, seiner Bewertung und Verwertung, in der Regel organisiert durch Verlage, der Erschließung, Aufbewahrung und Bereitstellung durch Bibliotheken und des Rezipierens durch die wissenschaftliche Gemeinschaft, aus deren Mitte dann wiederum Autorinnen und Autoren ein nächstes geistiges Werk, eine Publikation, schaffen. Die Beteiligten an diesem Kreislauf verfolgen sicher das gemeinsame Ziel, dem wissenschaftlichen Fortschritt zu dienen. Die Motivationen, sich an diesem Prozess zu beteiligen, sind jedoch unterschiedlich ausgeprägt und innerhalb der jeweiligen Gruppen bei weitem nicht einheitlich, wobei auf diesen Punkt jedoch nicht näher eingegangen werden soll.

Unabhängig davon kristallisieren sich für das wissenschaftliche Publizieren im Wesentlichen vier Funktionen heraus: Sie dient der Kommunikation, als Nachweisinstrument, zur Erlangung von Reputation sowie zur Generierung finanzieller Erträge.

- **Kommunikation:** Der Hauptzweck des wissenschaftlichen Publizierens besteht darin, die Kommunikation zwischen Wissenschaftlern über deren Forschungsfragen, -ergebnisse und -erkenntnisse in einer formalisierten Weise zu realisieren. Dabei soll dieser Informationsaustausch einschließlich der eigentlichen Inhalte in nachvollziehbarer Form für die weitere Verwendung und die eindeutige Bezugnahme dauerhaft zur Verfügung stehen. Selbstverständlich gibt es neben dem Publizieren andere Formen der wissenschaftlichen Kommunikation – etwa persönliche Gespräche oder wissenschaftliche Konferenzen. Abgesehen von der Nachvollziehbarkeit gewährleistet das wissenschaftliche Publizieren darüber hinaus jedoch die zeitliche und räumliche Entkoppelung der Kommunikation. Im Unterschied zu den synchronen Kommunikationsmöglichkeiten sind hierbei

sowohl die aktive als auch die passive Teilnahme an der Kommunikation mit geringeren Voraussetzungen verbunden. Die Kommunikation selbst findet dadurch auch in größerer Öffentlichkeit statt.

- **Nachweisinstrument:** Zweitens dient das Publizieren seit seinen Anfängen auch dazu, eigene Ansprüche an wissenschaftlichen Ergebnissen, Entdeckungen und Erkenntnissen wirksam kenntlich zu machen. Es ist damit Nachweisinstrument für Priorität und Originalität bzw. zum Nachweis von Plagiaten (vgl. [3]) und dient letztlich also zur Sicherung des Urheberrechts. Wissenschaftliches Publizieren steht in diesem Sinne für den wissenschaftlichen Wettbewerb, denn die Zweckbestimmung der Wissenschaft selbst liegt darin, Wissen als begründete Information zu produzieren, das bisher noch nicht publiziert worden ist (vgl. [4]). Die Vielzahl so genannter Prioritätsstreits (siehe [3]), die zwischen mitunter durchaus renommierten Wissenschaftlern ausgetragen wurden, verdeutlicht den Wert des wissenschaftlichen Publikationswesens für die Ermittlung und den Nachweis des Erstlingsrechts an wissenschaftlichen Erkenntnissen. Beispielsweise wurde Leibniz infolge seines Prioritätsstreits mit Newton über die Entwicklung der Infinitesimalrechnung durch die Royal Society nur deshalb des Plagiats für schuldig befunden, weil er seine bereits deutlich länger vorhandenen Erkenntnisse nicht veröffentlicht hatte.
- **Reputation:** In enger Verbindung mit der Sicherung und dem Nachweis der Priorität und des Urheberrechts ist eine weitere Funktion des wissenschaftlichen Publizierens zu sehen, die sich vorrangig aus Sicht der Wissenschaftler ergibt – der Nachweis über die eigenen erbrachten wissenschaftlichen Leistungen und letztlich die Steigerung der eigenen Reputation. Letztere hängt heute in den meisten Wissenschaftsdisziplinen von keinem Kriterium so stark ab wie von der Anzahl der eigenen Veröffentlichungen und deren formaler qualitativer Gewichtung, was treffend mit dem geflügelten Wort „Publish or Perish“ umschrieben wird.

Der damit einhergehende Anreiz aufseiten potentieller Autoren, möglichst viel zu publizieren, drängt nicht nur die ursprüngliche Zweckbestimmung des wissenschaftlichen Publizierens in den Hintergrund, er führt auch zu einer abnehmenden Effizienz des Publikationswesens – etwa durch die separate Veröffentlichung kleinster Fragmente (die so genannte „Salamiaktik“), wissenschaftlich wertloser oder banaler Arbeiten bis hin zum bewussten Fälschen von Ergebnissen. Der Druck, möglichst viele Publikationen vorweisen zu können, wirkt sich nicht nur auf das Verhalten der einzelnen Wissenschaftler aus, sondern beeinflusst auch ganze Institutionen in ihrem Handeln, die der auf der Publikationsanzahl basierenden Forschungsevaluation ebenfalls unterliegen. Letztlich bleibt festzustellen, dass der wesentliche Zweck des wissenschaftlichen Publizierens – die formalisierte Kommunikation untereinander – von einem anderen Zweck – der Steigerung der eigenen Reputation – überlagert wird, der zumindest teilweise die Merkmale eines Selbstzwecks besitzt<sup>2</sup>. Diese Problematik muss sich in besonderer Weise auf die Anforderungen an das wissenschaftliche Publizieren und insbesondere auf die Qualitätssicherung auswirken.

- **Finanzielle Erträge:** Als weiterer eher als eigennützig zu bezeichnender Zweck des Publizierens sei hier noch auf die Erlangung finanzieller Erträge aus der Veräußerung bzw. dem Verkauf von Subskriptionen für Publikationen verwiesen. Er spielt aus Sicht der Autoren als Motivation in vielen Wissenschaftsdisziplinen eine untergeordnete Rolle, da sie beispielsweise für Zeitschriftenartikel in der Regel keine Vergütungen erhalten. Dagegen werden beispielsweise im Falle von Lehrbüchern und anderen Monographien die Verfasser durchaus an den Erträgen der Vermarktung beteiligt.

<sup>2</sup> Bereits Ende der 80er Jahre des 20. Jahrhunderts stellte der Herausgeber des Journal of the American Chemical Society, Allen Bard, fest: „In many ways, publication no longer represents a way of communicating with your scientific peers, but a way to enhance your status and accumulate points for promotion of grants.“ [5]



Insbesondere für diejenigen Teilnehmer am Publikationsprozess – in der Regel die Verlage – die ihn organisieren und daran ein wirtschaftliches Interesse knüpfen, besteht selbstverständlich ein Hauptzweck oder gar der einzige Zweck darin, damit finanzielle Erträge zu erlangen. Deren anreizorientiertes Handeln beeinflusst das gesamte wissenschaftliche Publikationswesen in nicht unwesentlichem Umfang. Als Beispiel sei hier nur auf die so genannte Zeitschriftenkrise verwiesen, die durch eine monopolistische Preispolitik verschiedener Verlage hervorgerufen wurde und völlig neue Ansätze wissenschaftlicher Kommunikation hervorgerufen oder zumindest beflügelt hat und eine der Ursachen für das Entstehen der weltweiten Open-Access-Bewegung war.

## Anforderungen an das wissenschaftliche Publizieren

Die Anforderungen, die an das wissenschaftliche Publizieren zu richten sind, ergeben sich im Wesentlichen aus den im vorigen Abschnitt benannten Funktionen des Publizierens. Sie werden aus der Sicht der Wissenschaft insgesamt formuliert, die ihrerseits keinen Selbstzweck darstellt, sondern der gesamtgesellschaftlichen Funktion der Erschaffung, Bewahrung und Vermittlung von Wissen dient. Im Einzelnen können dabei folgende Aspekte unterschieden werden:

- **Zugänglichkeit:** Das wissenschaftliche Publizieren dient der Kommunikation zwischen Wissenschaftlern. Um diese Hauptfunktion erfüllen zu können, muss das Publikationswesen in seiner Gesamtheit für die Zielgruppe verfügbar und nutzbar sein – und zwar sowohl für Rezipienten als auch für potentielle Autoren. Der Zugang zu relevanten Publikationen wird in der Regel über die wissenschaftlichen Bibliotheken realisiert, die zur jeweiligen Institution gehören, an der die betreffenden Wissenschaftler tätig sind. Sie tragen sowohl für die Beschaffung und damit die grundsätzliche Verfügbarkeit als auch für Erschließung und Bereitstellung und damit die eigentliche Zugänglichmachung die Verantwortung.

Nachteile in Bezug auf die Zugänglichkeit haben Wissenschaftler, die nicht unmittelbar an einer Institution angestellt sind und auf deren Infrastruktur zurückgreifen können.

Zu thematisieren ist jedoch auch die Zugänglichkeit des Autors zum Publikationswesen an sich. Sie sollte so beschaffen sein, dass Wissenschaftlern grundsätzlich der Weg zur Publikation eigener Ergebnisse offensteht und sie nicht aufgrund von Kriterien, die nicht die wissenschaftliche Qualität der eigentlichen Arbeiten betreffen – etwa die Zugehörigkeit zu einer bestimmten Institution, Nationalität oder Religion, das Geschlecht, die eigenen finanziellen Möglichkeiten, die Interessen Dritter und dergleichen – dabei behindert oder gänzlich davon ausgeschlossen werden.

- **Nachhaltigkeit:** Das wissenschaftliche Publizieren – wie das Publizieren im Allgemeinen – soll grundsätzlich dazu geeignet sein, Werke so zu veröffentlichen, dass sie potentiell dauerhaft verfügbar bleiben. Das heißt, der gesamte Publikationsprozess, aber auch die Träger- bzw. Übertragungsmedien im Hinblick auf deren physikalische Eigenschaften sowie die Formate (Kodierung) müssen so beschaffen sein, dass sie die Nachhaltigkeit der Veröffentlichung erlauben. Dies beinhaltet insbesondere die Unveränderlichkeit der Publikationen oder zumindest die Nachweisbarkeit etwaiger Veränderungen. Das etablierte System der Bezugnahme auf bereits erschienene Publikationen mittels Zitationen basiert vor allem auf der Annahme, dass einmal veröffentlichte Werke zugänglich bleiben und sich fürderhin auch nicht mehr verändern (lassen).
- **Nachvollziehbarkeit:** Eine wesentliche Anforderung an das wissenschaftliche Publizieren ist die Nachvollziehbarkeit des gesamten Prozesses. Informationen zum zeitlichen Ablauf müssen in möglichst nachprüfbarer Form vorgehalten werden – vor allem das Datum der Einreichung eines Manuskriptes, das Datum der eigentlichen Veröffentlichung und die jeweils vorliegenden Manuskriptversionen. Zur Nachvollziehbarkeit zählt auch die eindeutige

und dauerhafte Identifizierbarkeit der einzelnen Publikationen, u. a. als Voraussetzung für die Zitierbarkeit und die Auffindbarkeit. In diesem Zusammenhang wird gerade in der letzten Zeit mit immer größerem Nachdruck die öffentliche Verfügbarkeit der so genannten Primärinformationen gefordert. Gemeint sind damit Informationen, die die Basis für den jeweils vorliegenden wissenschaftlichen Artikel gelegt haben, um auch sie hinsichtlich der Wahrhaftigkeit prüfen zu können. Diese Primärinformationen können sicher zitierte Artikel anderer Wissenschaftler sein, aber eben auch die Ergebnisse von Befragungen, Experimenten, Zeitreihen u. ä.

- **Authentizität:** Da sich im Gegensatz zu unmittelbaren Kommunikationsformen wie etwa dem persönlichen Gespräch die Kommunikationspartner beim wissenschaftlichen Publizieren nicht direkt begegnen, sondern lediglich mittelbar miteinander in Kontakt treten, sind die Glaubwürdigkeit und in diesem Zusammenhang das Vertrauen in die zwischen den eigentlichen Kommunikationspartnern – den Wissenschaftlern – agierende Instanz eine Grundvoraussetzung für das Publikationswesen. Diese Glaubwürdigkeit oder Authentizität bezieht sich insbesondere darauf, dass die Verfasser, die als Autoren einer Publikation genannt sind, tatsächlich deren Urheber sind und dass der in der Publikation wiedergegebene Inhalt in dieser Form tatsächlich von ihnen stammt. Als vertrauenswürdige Instanz fungiert hier in der Regel der Herausgeber (Editor), der letztlich auch die Verantwortung für die Korrektheit dieser (Meta-)Informationen trägt. Aber auch er muss sich zum Teil auf die Angaben der Verfasser, die ihm nicht in jedem Falle persönlich bekannt sind, verlassen. Dazu zählt beispielsweise auch die korrekte Anwendung der angegebenen wissenschaftlichen Methoden. Die Glaubwürdigkeit spielt aber nicht nur aus Sicht der Rezipienten und der Herausgeber eine Rolle. Auch die Verfasser leisten durch die Einreichung eines (zu diesem Zeitpunkt eben noch unveröffentlichten) Manuskriptes dem

Herausgeber gegenüber einen hohen Vertrauensvorschuss – vor allem in Bezug auf die rechtmäßige und faire Behandlung der eingereichten Arbeit, aber beispielsweise auch in Bezug auf die vertrauliche Behandlung des Manuskripts im Vorfeld der eigentlichen Veröffentlichung.

- **Qualitätssicherung:** Mit der Glaubwürdigkeit einer Publikation und des gesamten Publikationsverfahrens hängen auch die Kontrolle und Sicherung der inhaltlichen Qualität zusammen. Die Qualitätssicherung wird in der Regel durch einen gesonderten Prozess realisiert, der zwischen der Einreichung eines Manuskripts und dessen formeller Veröffentlichung stattfindet und unabhängig von dessen Autoren ist. Er soll sicherstellen, dass in einem Publikationsmedium – etwa in einer Zeitschrift – ein bestimmtes Level an Relevanz, Originalität und Qualität erreicht wird. Entscheidend für die Wirksamkeit der Qualitätssicherung und deren Glaubwürdigkeit ist eine vertrauenswürdige Instanz, die den Prozess organisiert – in der Regel die Herausgeber, die unabhängig von den Verfassern der infrage stehenden Manuskripte agieren (siehe dazu auch [6]).
- **Bewertung:** Die Qualitätssicherung erfolgt zeitlich vor der Veröffentlichung eines Werkes. Welche Wirkung und Bedeutung eine wissenschaftliche Publikation tatsächlich haben wird, lässt sich vor der eigentlichen Veröffentlichung jedoch allenfalls vermuten. Da eigene Publikationen und deren wissenschaftlicher Wert die Reputation der jeweiligen Autoren wesentlich beeinflussen (sollen), ist eine der Veröffentlichung nachgelagerte Messung von Indikatoren erforderlich, die Rückschlüsse auf Bedeutung und Wirkung einzelner Publikationen zulassen. Derartige Bewertungsverfahren, die sich beispielsweise auf Verkaufszahlen stützen können, hängen nicht notwendigerweise unmittelbar mit dem Publizieren im engeren Sinne zusammen. Bezieht man in den Publikationsprozess jedoch die gesamte Lebensdauer einer Publikation (bzw. der einzelnen Exemplare) einschließlich deren Nutzung und Zitierung ein, gehört diese

Anforderung durchaus hierher.

- **Geschwindigkeit:** Es liegt nahe, dass beim wissenschaftlichen Publizieren wegen der Mittelbarkeit, mit der die Kommunikation hierbei realisiert wird, wesentliche zeitliche Verzögerungen entstehen. Dies ergibt sich nicht nur aus technischen Randbedingungen der einzelnen Schritte (Einreichung, Satz, Druck, Auslieferung), sondern vor allem auch aus dem zeitlichen Aufwand, der aus den vorher genannten Anforderungen an das wissenschaftliche Publizieren (insbesondere Nachvollziehbarkeit und Qualitätssicherung) resultiert. Insofern läuft die Anforderung an eine möglichst zeitnahe Veröffentlichung eingereichter Manuskripte anderen Anforderungen entgegen.
- **Vollständigkeit:** Die Vollständigkeit als eine Anforderung an eine Publikation zu formulieren, verwundert auf den ersten Blick, weil sie sich aus inhaltlicher Sicht durch das wissenschaftliche Grundverständnis ergibt. Gemeint ist hier vielmehr die formale Vollständigkeit, die sich durch das Vorhandensein von Metadaten zur Beschreibung der Publikation ausdrückt. Diese Anforderung bekommt in der gegenwärtigen Zeit einen immer größeren Stellenwert, weil an die Stelle des Trägermediums Papier elektronische Speichermedien treten. Informationen, die sich zuvor durch die Papierveröffentlichung von selbst ergaben, wie das Format der Speicherung, der Umfang der Veröffentlichung, die innere Struktur des Dokuments u. Ä. sind zusätzlich zu erfassen, um den oben formulierten Anforderungen entsprechen zu können. Im Wesentlichen unterscheidet man zwischen beschreibenden Metadaten, verwaltungstechnischen Metadaten, strukturellen Metadaten, technischen Metadaten und Archivierungsmetadaten. (siehe dazu [7])

## Die Urheberschaft als Ausdruck von Verantwortlichkeit und Glaubwürdigkeit

Das gesamte Publikationswesen in der Wissenschaft ist darauf ausgerichtet, dass Wissenschaftler unterschiedlicher Institutionen insofern virtuell miteinander zusammenarbeiten können, als sie die Erkenntnisse ihrer unter Umständen weltweit verstreuten und potentiell zu unterschiedlichen Zeiten lebenden Fachkollegen in ihre eigenen Überlegungen und Forschungen mit einfließen lassen und sich darauf eindeutig beziehen können. Erkenntnisse, Entdeckungen und andere Resultate wissenschaftlichen Handelns müssen sich dabei nicht nur an ihrer prinzipiellen Güte, Korrektheit, Plausibilität oder anderen unabhängigen Qualitätskriterien messen lassen. Ihr Wert hängt auch davon ab, wie sie zu bereits bekannten – in der Regel also veröffentlichten – Ergebnissen zu demselben Thema in Beziehung stehen und insbesondere, ob die infrage stehenden Erkenntnisse bzw. Entdeckungen bereits anderweitig publiziert wurden. Insofern ist das wissenschaftliche Publizieren nicht nur die Voraussetzung dafür, dass wissenschaftliche Erkenntnisse unabhängig von den Personen, die sie ursprünglich hervorgebracht haben, erhalten bleiben, sondern es dient auch dazu, dass sie weiterentwickelt und als Grundlage weiterer Forschung und Wissenschaft verwendet werden können. Bei einer gesamtgesellschaftlichen, eher ergebnisorientierten Betrachtung ist dabei vor allem der wissenschaftliche Inhalt der Publikationen von Interesse, wohingegen die Notwendigkeit, diesen Inhalt mit dessen Urhebern in Verbindung zu bringen, eher in den Hintergrund tritt. Damit das System in sich funktioniert, spielt jedoch die Information, wer eine Veröffentlichung verfasst hat, eine wesentliche Rolle. Dies hat zum einen mit der Reputation und dem Nachweis der Leistungen der betreffenden Autoren zu tun und betrifft somit das gesamte Anerkennungssystem in der Wissenschaft. Andererseits wird durch die explizite Nennung der Urheberschaft einer Publikation und die damit verbundene Zurückführung der

darin enthaltenen Informationen auf deren Autoren insgesamt ein Netz an Verantwortlichkeit und Glaubwürdigkeit aufgebaut. Beide Aspekte messen den Personen, die an der Entstehung wissenschaftlicher Ergebnisse maßgeblich beteiligt sind, eine zentrale Bedeutung zu – zum einen in Form von Anerkennung, zum anderen in Form von Verantwortung. Dass dieses grundsätzliche Publikationsparadigma durchaus keine Allgemeingültigkeit besitzt, zeigt unter anderem das Beispiel der Wikipedia, in der die Urheberschaft einzelner Artikel hinter deren Inhalt fast vollständig zurücktritt. Auf der anderen Seite ist die Bedeutung, die der Rolle der Urheberschaft einer Publikation im traditionellen Publikationswesen allgemein beigemessen wird, wiederum dazu geeignet, durch missbräuchliche Angaben persönliche, interessegeleitete Ziele zu verfolgen. Beispielsweise sei hier nur auf die so genannten unethischen Autorenschaften bei Publikationen verwiesen, die zuweilen auch als Ehrenautorenschaften bezeichnet werden. Als Autoren solcher Veröffentlichungen werden neben den eigentlichen Verfassern andere Personen genannt, die damit einerseits die Liste der „eigenen“ Publikationen anreichern und andererseits durch ihre eigene Bekanntheit die Glaubwürdigkeit und damit die Akzeptanz der Publikation in der Wissenschaftswelt unabhängig von deren inhaltlicher Qualität erhöhen.

## Verlage und Herausgeber – Vermittler im Publikationswesen

Das wissenschaftliche Publikationswesen ist seit seinen Anfängen Ausdruck der Selbstorganisation der Wissenschaft. Neben den Autoren sind in der Regel auch die Herausgeber von Zeitschriften, Sammelbänden, Schriftenreihen und dergleichen aktive Wissenschaftler und damit selbst Teil der Community der jeweiligen Fachdisziplin. Sie bilden einerseits die vertrauenswürdige Instanz, die einer Publikation zu Glaubwürdigkeit verhilft. Andererseits obliegt ihnen die Organisation des Publikationsprozesses, in dessen Kern vor allem die Begutachtung steht, mit der die Qualitätssiche-

rung der wissenschaftlichen Inhalte sichergestellt werden soll. Während Autoren, wissenschaftliche Herausgeber und Gutachter in der Regel aktive Wissenschaftler sind, können organisatorische, technische und finanzielle Aspekte des Publikationsprozesses grundsätzlich durchaus auch durch Dritte übernommen werden. Aus dieser Einsicht in die Vorzüge einer arbeitsteiligen Realisierung des wissenschaftlichen Publizierens entstanden Verlage, die zunächst in der Rolle von Dienstleistern für die Wissenschaft auftraten. Neben der Organisation des gesamten Publikationsprozesses besitzen Verlage typischerweise höhere Kompetenzen in den Bereichen Marketing und Vertrieb und können damit für eine möglichst große Verbreitung von Publikationen sorgen. Als wirtschaftlich handelnde Einheiten bestand seitens der Verlage selbstverständlich von Anfang an eine gewisse Gewinnorientierung. Dennoch ist die weit verbreitete Arbeitsteilung zwischen der inhaltlichen Arbeit von Wissenschaftlern als Autoren, Herausgeber, Gutachter und Rezipienten und allen anderen Bereichen des Publikationswesens vor allem im Hinblick auf eine Entlastung des Wissenschaftsbetriebes von nichtwissenschaftlichen Tätigkeiten durchaus zweckmäßig.

Mit der zunehmenden Etablierung monopolartiger Strukturen und einer entsprechend aggressiven Preispolitik vor allem bei Zeitschriftenverlagen im STM-Bereich<sup>3</sup> geriet diese auf beiderseitigem Nutzen bestehende Beziehung zwischen Wissenschaft und Verlagen jedoch zunehmend in ein Missverhältnis. Weil die Bibliotheksetats mit den übermäßigen Preiserhöhungen und der bis heute andauernden Steigerung der Anzahl publizierter Zeitschriften nicht mithalten können, werden einerseits vermehrt Titel abbestellt bzw. können erst gar nicht bestellt werden. Zum anderen werden alternative Publikationsmodelle entwickelt, die beispielsweise auf Open-Access-Geschäftsmodellen (siehe [8]) beruhen und/oder Verlage aus der Publikationskette grundsätzlich ausklammern.

<sup>3</sup> STM = Science, Technology, Medicine

## Perspektiven des wissenschaftlichen Publizierens

Das CMS-Journal bietet nicht den Platz, den sich vollziehenden Veränderungsprozess beim wissenschaftlichen Publizieren ausführlich und im Detail zu beschreiben (vgl. dazu [9]) und die Konsequenzen in ihrer Gesamtheit aufzuzeigen. Im Wesentlichen zeichnen sich jedoch drei Entwicklungen ab, die im Folgenden kurz skizziert werden:

1. *Das wissenschaftliche Publizieren ist im Wandel, weil technische Schranken der Vervielfältigung und Verbreitung gefallen sind und somit ein freier uneingeschränkter Zugriff auf das Wissen der Welt möglich wird.*

Die Vervielfältigung wissenschaftlicher Arbeiten war über Jahrhunderte ein mehr oder weniger großes technisches Problem. Bevor es die „bewegliche Letter“ in Europa gab, war das Abschreiben von Büchern die einzige und in seinem Umfang sehr begrenzte Möglichkeit, Kopien zu erzeugen. Zwangsläufig waren damit auch der Verbreitung sehr enge Grenzen gesetzt. Der technische Fortschritt des Buchdruckes und des entstandenen Versandhandels verringerten diese Barrieren zwar wesentlich, hoben sie jedoch in keiner Weise auf. Erst durch die Überführung des Publikationsprozesses in die digitale Welt, mit der unter anderem verlustfreie Kopien in unbegrenzter Anzahl und zu verschwindend geringen Kosten erzeugt werden können – beispielsweise auch durch den einzelnen Rezipienten selbst – ergibt sich eine völlig neue Qualität an Verbreitungsmöglichkeiten. Der weltweite, unmittelbare und durchaus auch freie Zugriff auf das dergestalt vorhandene Wissen der Welt ist damit aus technischer Sicht erstmals prinzipiell möglich. Im weiteren Verlauf des vorliegenden Heftes wird zu zeigen sein, welche Möglichkeiten und Grenzen bei der Umsetzung der Prinzipien des Open-Access-Publizierens gesehen werden.

2. *Die Möglichkeiten des elektronischen Publizierens vereinfachen den technischen Vorgang des Erstellens einer Publikation erheblich, erlauben damit eine zeitnahe*

Veröffentlichung und lassen eine unmittelbare weltweite Verbreitung realistisch erscheinen.

Der Arbeitsplatz eines Wissenschaftlers der heutigen Zeit ist durch den Computer geprägt. Jede Veröffentlichung einschließlich eines Großteils der dazu notwendigen Vorarbeiten entsteht auf einem solchen Computer, wird damit bearbeitet und findet dort seinen „ersten“ Speicherplatz. Der einzelne Autor kann somit sowohl über den Zeitpunkt der Verbreitung als auch über den Verbreitungsgrad entscheiden. Er kann durch den Austausch per E-Mail oder die Nutzung von Preprint-Servern zu möglichen Diskussionen vor einer umfassenden Veröffentlichung anregen und somit zusätzliche Möglichkeiten der Qualitätsverbesserung nutzen.

3. *Durch die Möglichkeiten der digitalen Repräsentation und Verarbeitung, die vielfältigen Möglichkeiten der Speicherung auf elektronischen Medien und die Bereitstellung und Verbreitung über das Internet werden die Publikationsformen der Zukunft nicht mehr ausschließlich text- und grafikorientiert sein, sondern ebenso durch Audio- und Videobestandteile bestimmt werden.*

Die wissenschaftliche Publikation der Vergangenheit war gleichzusetzen mit einer textorientierten Veröffentlichung, die teilweise durch entsprechende Grafiken illustriert wurde. Für diese Art der Darstellung war der Mensch in der Lage, die eingangs formulierten Anforderungen an eine Publikation sicherzustellen und die resultierenden Exemplare über einen beachtenswerten Zeitraum unverfälscht aufzubewahren. Durch die Vielzahl der entwickelten Regeln des Umgangs mit einer wissenschaftlichen Publikation (Peer Review, Erschließungsregeln, Katalognachweis, Zitationsstile, Nutzungsstatistiken u. v. a.), die in der Zwischenzeit zum „guten Ton“ des Publizierens gehören, war ein in sich stabiles System entstanden. Die Grundlage dieses Systems bildet das Trägermedium Papier, dessen Herstellung, Verarbeitung, Aufbewahrung und vor allem Haltbarkeit einer ständigen Verbesserung unterzogen wurde. Für wissenschaftliche Phäno-

mene, die sich nicht auf Papier darstellen lassen, bestand dagegen keine adäquate Verbreitungsmöglichkeit, was Beschreibung und Forschung in den entsprechenden Bereichen wesentlich begrenzte. Diese Einschränkung lässt sich durch die Nutzung digitaler Datenformate und multimedialer Darstellungsformen überwinden. Die digitale Speicherung von Bewegtbild und Ton macht es möglich, dass diese Darstellungsarten neben Text und Grafik gleichberechtigt aufgezeichnet, aufbewahrt und wieder aufgefunden werden können. Man wird in den kommenden Jahren beobachten können, wie sich diese Publikationsformen entwickeln und welche Regeln sich heraus bilden, um den Anforderungen an eine wissenschaftliche Publikation gerecht zu werden.

Parallel zu diesen technischen Neuerungen müssen und werden sich neue organisatorische Randbedingungen entwickeln, die bei den Beteiligten des Publikationsprozesses veränderte Verhaltensweisen erzeugen. In den nachfolgenden Beiträgen dieser Ausgabe des cms-journals wird davon einiges zu lesen sein. Mit diesem Veränderungsprozess steht die wissenschaftliche Community noch ganz am Anfang. Die angesprochenen technischen Neuerungen sind teilweise noch nicht ausgewogen und bedürfen weiterer Anstrengungen, um nicht hinter dem gegenwärtigen Standard des wissenschaftlichen Publizierens zurückzubleiben. Andere Geschäftsmodelle für das Publizieren stehen zur Diskussion, um zum einen die inhaltliche Qualität des Produkts zu sichern und zum anderen den freien Zugriff auf das Wissen der Welt zu ermöglichen. Die technischen Hilfestellungen für Autorinnen und Autoren, aber auch für Bibliotheken und Rezipienten sind bei weitem nicht ausgeschöpft und vor allem in Bezug auf die Ausnutzung der sich ergebenden Vorteile der neuen technischen Möglichkeiten noch nicht optimiert.

## Literatur

- [1] RIEHM, U.; BÖHLE, K.; WINGERT, B.: *Elektronisches Publizieren*. In: Kühlen, R.; Seeger, T.; Strauch, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. München: K.G. Saur, 2004, S. 549–558.
- [2] ZUCKERMAN, H.; MERTON, R. K.: *Institutionalized Patterns of Evaluation in Science*. In: Storer, N. W. (Hrsg.): *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, London: The University of Chicago Press, 1973, S. 460–496.
- [3] MERTON, R. K.: *Priorities in Scientific Discovery: A Chapter in the Sociology of Science*. *American Sociological Review* 22 (6), 1957, S. 635–659.
- [4] UMSTÄTTER, W.: *Was ist und was kann eine wissenschaftliche Zeitschrift heute und morgen leisten*. In: Parthey, H.; Umstätter, W. (Hrsg.): *Wissenschaftliche Zeitschrift und Digitale Bibliothek: Wissenschaftsforschung Jahrbuch 2002, 2002*, S. 143–166.
- [5] HAMILTON, D. P.: *Publishing by – and for? – the Numbers*. *Science* 250 (4986), 1991, S. 1331–1332.
- [6] MÜLLER, U.: *Peer Review bei Open-Access-Zeitschriften*. *cms-journal* 32, 2009, S. 33–37.
- [7] DOBRATZ, S.: *Grundfragen der digitalen Langzeitarchivierung für den edoc-Server*. *cms-journal* 32, 2009, S. 93–98.
- [8] GRADMANN, S.: *Publizieren im Open-Access-Modell*. *cms-journal* 32, 2009, S. 20–23.
- [9] SCHIRMBACHER, P.: *Die neue Kultur des elektronischen Publizierens*. In: Nielsen, E. K.; Saur, K. G.; Ceynowa, K. (Hrsg.): *Die innovative Bibliothek – Elmar Mittler zum 65. Geburtstag*. München: K.G. Saur, 2005, S. 107–120.

# Entwicklungsstufen des wissenschaftlichen Publizierens

Diese Eckdaten sind als Auswahl zu verstehen und haben nicht den Anspruch der Vollständigkeit.

## Erste Meilensteine des Publizierens

- 3000 Entwicklung der Schrift (Mesopotamien)
- 105 Erfindung des Papiers (China)
- 1040 Erfindung des Buchdrucks (China)
- um 1450 Erfindung des Buchdrucks mit beweglichen Lettern (Europa: Gutenberg)
- 1665 erste wissenschaftliche Zeitschriften

## Entwicklung des Elektronischen Publizierens

- 1980 erste elektronische Zeitschrift „Mental Workload“
- 1990 erste Webseite durch Tim Berners Lee
- 1991 Preprint- und Dissertationsserver (zunächst für Physik: frei zugänglicher ArXiv-Server)
- 1993 Freigabe des WWW durch CERN
- 1994 „Subversive Proposal“: Prinzip des Self-Archiving (Steven Harnad)
- 1997 Gründung von SPARC (Scholarly Publishing and Academic Resources Coalition)
- 1999 Gründung der OAI (Open Archives Initiative), PMH (Protocol for Metadata Harvesting)
- 2001 Gründung von Public Library of Science als Open-Access-Verlag

## Etablierung von Open Access

- 2001 Konferenz zu Open Access in Budapest: BOAI (Budapester Open Access Initiative), anschließend weltweiter OA-Aufruf als „Geburtsurkunde“ der OA-Bewegung
- 2003 Konferenz in Bethesda: Definition von „offen zugänglicher Veröffentlichung“, verstärkte Einbindung der Autoren
- 2003 „Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“: erweiterte Zielsetzung, Unterzeichnung der großen deutschen Forschungseinrichtungen
- 2006 Verabschiedung der „Open-Access-Erklärung der Humboldt-Universität zu Berlin“

## Open Access in der Gegenwart

- weltweit ca. 1300 OA-Repositories in OpenDOAR (Open Directory of Open Access Repositories) bzw. in ROAR (Registry of Open Access Repositories) registriert, davon ca. 140 in Deutschland
- weltweit ca. 4000 OA-Journals (mit Peer Review-Verfahren) im DOAJ (Directory of Open Access Journals) registriert, davon ca. 150 aus Deutschland
- übergreifende Dienste:
  - DRIVER (Digital Repository Infrastructure Vision for European Research)
  - OA-Netzwerk (Open-Access-Netzwerk)
  - OAIster (Suche im Bestand der bei OAI registrierten Anbieter)
  - BASE (Bielefeld Academic Search Engine: Suche im Bestand der bei OAI registrierten Anbieter und weitere Quellen)
  - SHERPA/RO-MEO (Securing a Hybrid Environment for Research Preservation and Access / Rights METadata for Open Archiving)
  - OA-Policies: Was gestatten deutsche Verlage ihren Autoren?
- weiterführende Informationen:
  - DINI e.V. (Deutsche Initiative für Netzwerkinformation), [www.dini.de](http://www.dini.de)
  - IPOA (Informationsplattform open-access.net)



# Möglichkeiten und gegenwärtige Grenzen des elektronischen Publizierens

Peter Schirmbacher | schirmbacher@cms.hu-berlin.de

## Einordnung des Artikels

„Das wissenschaftliche Publizieren befindet sich im Wandel“ ist eine in den letzten Jahren häufig in den Mittelpunkt gestellte Behauptung (siehe u. a. [1]). Mit dem Beitrag „Das wissenschaftliche Publizieren – Stand und Perspektiven“ [2] sollte dazu eine entsprechende Einstimmung gegeben werden. Ziel dieses Beitrages ist es, etwas mehr die technologische Komponente dieses Prozesses zu diskutieren. Dabei sollen zum einen die zweifelsfrei vorhandenen Vorteile des elektronischen Publizierens und zum anderen die gegenwärtig noch vorhandenen Probleme dargestellt werden. Der Beitrag richtet sich in erster Linie an die Fachkolleginnen und -kollegen auf dem Gebiet des elektronischen Publizierens, aber auch an interessierte Autorinnen und Autoren.

Das wissenschaftliche elektronische Publizieren hat im Vergleich zum wissenschaftlichen Publizieren überhaupt eine extrem kurze Entwicklungsgeschichte zu verzeichnen. Voraussetzung für diese Art des Publizierens ist zum einen die Möglichkeit einer digitalen Darstellung der Ergebnisse der wissenschaftlichen Arbeit, um auf diese Weise eine neue Form der Vervielfältigung zu ermöglichen, und zum anderen die weltweite Computervernetzung als neuer Distributionskanal wissenschaftlicher Arbeiten. Aufgrund der unterschiedlichen Entwicklungslinien auf dem Weg zur Digitalisierung von wissenschaftlichen Texten bzw. der Computervernetzung zum Zweck der Datenübertragung ist es nicht ohne weiteres möglich, eine konkrete Zeit für den „Beginn“ des elektronischen Publi-

zierens zu bestimmen. Es ist sicher nicht falsch, diesen Zeitpunkt in die zweite Hälfte des vergangenen Jahrhunderts zu legen. Peter Suber ist mit seiner „Open Access Timeline“ [3] etwas konkreter. Er unterscheidet zwischen den Jahren vor 1990 und den dann „Schlag auf Schlag“ folgenden Aktivitäten der 90er Jahre. Das Ziel seiner Darstellung besteht jedoch nicht vordergründig darin, eine Historie des elektronischen Publizierens zu verfassen, sondern vielmehr die Geschichte der Open-Access-Aktivitäten aufzulisten.

Das für das elektronische Publizieren bedeutungsvollste Datum in seinem geschichtlichen Überblick ist sicher der 13. November 1990, an dem Tim Bernhards Lee die erste Webseite programmierte. Damit hat er letztlich den Grundstein für eine neue Veröffentlichungsform gelegt, die nicht nur den Computerspezialisten zugänglich ist, sondern im weiteren Verlauf der Entwicklung des WWW auch der breiten Öffentlichkeit. Das Datum ist auch deshalb als ein gutes Startsignal für eine neue Etappe in der wissenschaftlichen Kommunikation anzuerkennen, weil es das erklärte Ziel von Tim Bernhards Lee war, eine neue Kommunikationsform für den Austausch wissenschaftlicher Ergebnisse zwischen den Hochenergiephysikern zu schaffen, was ihm zweifelsfrei gelungen ist.

Nichtsdestotrotz liegen vor diesem Datum richtungsweisende Entwicklungen, die das elektronische Publizieren geprägt haben. Da ist zum einen das Transmission Control Protocol/Internet Protocol (TCP/IP) (1983) als Grundlage für die Kommunikation über das Internet zu nennen, zum zweiten die Schaffung der Standard

*Der Artikel konzentriert sich vorrangig auf technologische Komponenten des elektronischen Publizierens. Dabei stehen zum einen die neuen Möglichkeiten, die sich durch die Anwendung der digitalen Medien für das Publizieren ergeben und zum anderen die damit zur gegenwärtigen Zeit noch verbundenen Probleme im Mittelpunkt. Insbesondere werden die Vorteile des elektronischen Publizierens dargelegt und dabei wird auf die sich ändernden Anforderungen auch beim Erstellen von Texten durch Autoren eingegangen. Im gleichen Maße als wichtig wird jedoch auch das Herausarbeiten von Akzeptanzbarrieren erachtet, um damit Wege zur Beseitigung von Hemmnissen aufzuzeigen.*

*Elektronisches Publizieren, digitale Dokumente, Retrieval, strukturiertes Schreiben, multimediale Dokumente, Langfristarchivierung*

Generalized Markup Language (SGML) (seit 1986 eine ISO-Norm) als der Basis-sprache für alle später entwickelten Auszeichnungssprachen und zum weiteren die vielen Aktivitäten zur Verzeichnung von wissenschaftlichen Abhandlungen in elektronisch verfügbaren Katalogen, die zwar nur sehr selten auf einen elektronisch verfügbaren Volltext, wie wir es heute schon beinahe selbstverständlich erwarten, verwiesen, aber doch eine neue Qualität der Informationsrecherche ermöglichen.

## Was versteht man unter elektronischem Publizieren?

Erstaunlicherweise gibt es verhältnismäßig wenige Versuche, das elektronische Publizieren zu definieren. Uwe Müller [4] greift in seiner Arbeit auf Riehm et al. [5] zurück, die formulieren: "Elektronisches Publizieren umfasst öffentliche und zeitpunktunabhängige Formen indirekter Kommunikation über anerkannte Kanäle von derzeit überwiegend textlichen und grafischen Informationen in digitaler Form, wobei computerspezifische und multimediale Möglichkeiten zunehmend zum Einsatz kommen. Für die Nutzung elektronischer Publikationen wird Hard- und Software vorausgesetzt." Er führt dann aus: "Unter einer elektronischen Publikation ist in Anlehnung an die im Abschnitt 2.1.1<sup>1</sup> gegebene Definition eine Veröffentlichung zu verstehen, die sich von anderen Veröffentlichungen dadurch unterscheidet, dass sie in einer besonderen Repräsentation vorliegt – nämlich digital und auf einem elektronischen Medium." [4]

Deutlich wird an dieser Herangehensweise, dass man, wie es auch Müller (ebenda S. 6) tut, unterscheiden muss zwischen dem geistigen Werk als Resultat des elektronischen Publizierens und dem Vorgang der Schaffung einer elektronischen Publikation.

Unter Nutzung der aufgeführten Quellen wird auf folgende Definitionen orientiert:

<sup>1</sup> siehe [4] Seite 14: „Unter einer Publikation bzw. einer Veröffentlichung wird im Allgemeinen ein öffentlich verfügbares Werk in einer spezifischen Repräsentation verstanden.“

Unter einer elektronischen Publikation versteht man ein (in sich abgeschlossenes) öffentlich verfügbares Werk, das in digitaler Form auf einem elektronischen Medium vorliegt. Mit dem Begriff des elektronischen Publizierens im engeren Sinne bezeichnet man (dann zwangsläufig) den Vorgang der Erstellung einer elektronischen Publikation.

Da diese Definition eher einengt, wie im weiteren Text zu belegen ist, wird auf eine breitere Auslegung orientiert.

Unter dem elektronischen Publizieren im weiteren Sinne wird der gesamte Prozess der Erstellung, Verarbeitung, Speicherung und öffentlichen Bereitstellung einer elektronischen Publikation verstanden.

Deutlich sollte zunächst geworden sein, dass eine elektronische wissenschaftliche Publikation als eine Form einer wissenschaftlichen Publikation aufzufassen ist, die im Unterschied zu den bisher im Fokus stehenden Publikationen nicht analog, sondern digital vorliegt. Während bei einer analogen Publikation eine unmittelbare Bindung an das Trägermedium „Papier“ Voraussetzung ist, benötigt der Rezipient einer elektronischen Publikation „andere Hilfsmittel“, gemeinhin Hard- und Software, zur Wahrnehmung des geistigen Werkes. Riehm et al. [5] betonen dann, dass allgemein „anerkannte Kanäle“ für die Kommunikation die Grundlage für eine elektronische Publikation sind. Sie meinen damit, dass es zwingend eine Kompatibilität zwischen der eingesetzten Hard- und Software des die Publikation erzeugenden Autors und der Hard- und Software des Rezipienten geben muss, was außer Frage steht. Vergleicht man diese Phänomene genauer, so benötigt die elektronische Publikation neben dem elektronischen Speichermedium, auf dem die Publikation aufbewahrt ist, ein entsprechendes Gerät zur Sichtbarmachung. Das ist in der heutigen Zeit in der überwiegenden Zahl der Fälle der Computer mit seinem Bildschirm. Schon heute deuten sich jedoch Alternativen an, wenn man an die e-Book-Reader oder das so genannte elektronische Papier denkt.

Formal könnte man sich nun auf die Position zurückziehen, dass das elektronische Publizieren lediglich den Vorgang der Erstellung dieser elektronischen Publikation umfasst. Das würde bedeuten, dass sich der Untersuchungsgegenstand beim elektronischen Publizieren lediglich als der Prozess des „Einbringens“ des geistigen Werkes in digitaler Form auf ein entsprechendes Speichermedium darstellt. Da eine solche Definition prinzipiell sicher nicht falsch ist, wird sie jedoch als „im engeren Sinne“ bezeichnet, weil einer der Grundanforderungen an eine Publikation – die öffentliche Sichtbarmachung – damit nicht entsprochen werden würde.

In Erweiterung dazu schließt nun „das elektronische Publizieren im weiteren Sinne“ den gesamten, traditionellerweise als Publikationskreislauf bezeichneten Vorgang von der Erstellung bis zum Rezipieren ein. Durch das elektronische Publizieren erfährt dieser Kreislauf eine Vielzahl von Veränderungen aus technologischer Sicht, aber nicht nur aus diesem Blickwinkel, wie Gradmann in [6] unterstreicht.

Das elektronische Publizieren befindet sich am Anfang seiner Entwicklung, sodass man die „reine“ Form, bei der der gesamte Publikationsprozess ohne Medienbruch elektronisch verläuft, heute noch relativ selten antrifft. Entscheidend für die Benutzung des Begriffes „Elektronisches Publizieren“ sollte deshalb sein, dass das Zielausgabeformat ein digitales Format darstellt und die hauptsächliche Verbreitung über das Internet beabsichtigt ist. In diesem Sinne ist das vorliegende Heft des cms-journals kein Beispiel für das elektronische Publizieren, weil sich, wie sicher leicht zu erkennen ist, das Ausgabeformat an der Papierdarstellung ausrichtet. Zweifelsfrei finden jedoch schon eine Reihe von Komponenten des elektronischen Publizierens Berücksichtigung, wie andererseits an der Webdarstellung auf dem edoc-Server deutlich wird (siehe dazu [7]).



## Vorteile des elektronischen Publizierens

Bei der Diskussion der Vorteile des elektronischen Publizierens sind drei Kategorien zu unterscheiden: Zum Ersten wird der Publikationsprozess in seiner Gesamtheit effizienter gestaltet, zum Zweiten werden vorhandene Merkmale des Publizierens auf eine neue Qualitätsstufe gehoben und zum Dritten kommen neue Möglichkeiten hinzu, die mit der bisherigen Technologie nur teilweise oder gar nicht umsetzbar waren.

### Effizientere Gestaltung des Publikationsprozesses

Der bisherige Publikationsprozess war durch einen oder mehrere Medienbrüche geprägt und dadurch in keiner Weise effizient. Gemeint ist damit, dass die Publikation in Abhängigkeit vom Zeitalter, in dem sie entstand, zunächst handschriftlich oder mit der Schreibmaschine verfasst in Papierform vorlag. Daraus wurde eine Druckvorlage erarbeitet, indem der Text abgeschrieben, neu gesetzt bzw. verfilmt wurde. Es entstand eine Kopie auf einem anderen Medium, was wiederum dazu diente, ein Buch oder eine Zeitschrift nun wieder auf dem Medium Papier zu drucken. Beim elektronischen Publizieren in seiner ausgeprägtesten Form gibt es keinen Medienbruch, sondern der wissenschaftliche Beitrag entsteht auf einem Computer, wird dort entsprechend gespeichert und liegt somit digital vor. Zur Vorbereitung der Veröffentlichung bleibt er in dieser digitalen Form, wird möglicherweise bearbeitet, einem vorgegebenen Layout angepasst oder mit anderen Texten verbunden. Alle diese Vorgänge erfolgen unter Nutzung von digitalen Medien, wodurch zum einen eine verlustfreie (oder zumindest verlustarme) Kopie erzeugbar ist und dies zum anderen entweder automatisch oder mit nur geringem manuellem Aufwand erfolgt. Diese Technologie stellt einen enormen Produktivitätsgewinn dar, sodass die Aufwendungen deutlich geringer sind als bei der herkömmlichen Herangehensweise. Vielfach ist dieser ideale Prozessablauf in den 18 Jahren seit der Existenz des elektronischen Publizierens noch

nicht erreicht. Es gibt jedoch keine prinzipiellen Hindernisse, sondern nur Unzulänglichkeiten in der eingesetzten Technologie. In vielen Fällen fehlen etablierte Standards oder zumindest die internationale Verständigung auf gemeinsame Regelungen, die die Kompatibilität zwischen konkurrierenden Lösungen gewährleistet. All das ist Gegenstand der Forschung, sodass man davon ausgehen kann, dass in einem überschaubaren Zeitraum diese „Kinderkrankheiten“ überwunden sind. Einige der potentiellen Lösungsansätze finden sich in den weiteren Beiträgen dieses cms-journals.

Durch diese effizientere Technologie ergibt sich somit die prinzipielle Möglichkeit der Beschleunigung des Publikationsprozesses bei insgesamt geringerem Kostenaufwand.

### Qualitative Verbesserung der Technologie des Publizierens

Im Einführungsartikel des cms-journals [2] wurde hervorgehoben, dass sich durch die digitale Darstellung und die Computervernetzung die prinzipielle Möglichkeit der Verfügbarkeit der wissenschaftlichen Publikation überall auf der Welt ergibt. Es existieren dafür keine technischen Schranken mehr und auch die technisch verursachten Kosten für die notwendige Hard- und Software sind in der Zwischenzeit in vielen Teilen der Welt beherrschbar geworden. Der Aufwand zur Verbreitung im Sinne der Vervielfältigung (Kopienproduktion, Auflagenhöhe) und der Verfügbarkeit vor Ort (Bereitstellung durch eine Bibliothek oder den Buchhandel) haben eine völlig neue Qualität erreicht. Nicht mehr die an der Wirtschaftlichkeit ausgerichtete Auflagenhöhe einer Zeitschrift oder einer Monographie bestimmen die Zugriffsmöglichkeiten auf das geistige Werk, sondern allein der Zugang des Lesenden zu entsprechender Hard- und Software, die entsprechend mit dem Internet verbunden ist. Gelingt es, die Gedanken der Open-Access-Initiative zu wissenschaftlicher Information tatsächlich umzusetzen und damit von wirtschaftlichen Interessen zu befreien, dann ist das wissenschaftliche Publizieren in eine neue Dimension vorgestoßen.

Ähnlich gravierend kann man die sich abzeichnenden Verbesserungsmöglichkeiten für das Retrieval sehen. Das Erschließen wissenschaftlicher Werke als Basis für die Informationssuche ist nahezu so alt wie die wissenschaftlichen Bibliotheken selbst. Die Systeme wurden verfeinert und damit die Wahrscheinlichkeit, die für die wissenschaftliche Arbeit adäquate Literatur zu finden, immer größer. Es gab und gibt die so genannte Fernleihe, die den Zugriff auf Literatur ermöglichte, die nicht am jeweiligen Standort des Wissenschaftlers verfügbar war oder ist. In ähnlicher Weise tragen die in Deutschland etablierten Verbundsysteme mit den angeschlossenen Katalogen zu dieser positiven Entwicklung bei. Trotzdem stellt im Verhältnis dazu die Suchmaschinenteknologie eine völlig neue Qualität dar.

Diese Aussage wird gegenwärtig möglicherweise nicht von allen Lesern geteilt, noch dazu, wenn diese neue Qualität als eine bessere Qualität interpretiert werden soll. Die Hauptargumente der Gegner dieser These sind zum einen das Chaos, das die unüberschaubare Zahl von Dokumenten im Netz darstellt, und zum zweiten die Undurchdringlichkeit des Netzes für Suchmaschinen. Unbestritten ist es ein Problem, dass heute jeder ohne jede Kontrolle ins Netz stellen kann, was er möchte. Dieses Phänomen macht vor dem wissenschaftlichen Sektor keinen Halt, hat andererseits aber auch den Vorteil, dass es keine Zensur gibt, wie das Peer Review bei wissenschaftlichen Veröffentlichungen von einigen Verfechtern des freien Zugangs auch bezeichnet wird. Sicher muss es darum gehen, sich an dieser Stelle international auf Regeln und – wenn möglich – Standards zu verständigen, wozu z. B. eine solche Aktivität wie das DINI-Zertifikat (siehe [8]) einen Beitrag aus Deutschland darstellt.

Wenn man von der Undurchdringlichkeit des Netzes für Suchmaschinen spricht, dann wird gemeinhin das so genannte Deep Web gemeint, also der Teil des Netzes, der hinter Firewalls, zugriffsgeschützten Bereichen oder in speziellen Systemen, wie es zum Beispiel klassische Bibliothekssysteme darstellen, versteckt ist. Zu dieser Kategorie zählen gegenwärtig natürlich auch die

Server der Verlage, sodass der Zugriff für Suchmaschinen auf die dort herausgegebenen wissenschaftlichen Zeitschriften nicht möglich wird. Das sind Probleme, mit denen sich die Community auseinandersetzen muss. Sie sind „hausgemacht“, womit zum Ausdruck gebracht werden soll, dass diese Hindernisse nicht technisch bedingt sind, sondern ein Überbleibsel des bisherigen gesellschaftlichen Handelns. Leider ist davon auszugehen, dass sich diese neue Kultur einer barrierefreien Bereitstellung wissenschaftlicher Publikationen in unmittelbar nächster Zeit nicht durchsetzen wird. Käme es jedoch zu einem uneingeschränkten Zugriff auf alle Ressourcen des Netzes, so wären die Suchmaschinen den klassischen Recherchesystemen überlegen. Sie wären ihnen überlegen, weil sie deutlich schneller mit Ergebnissen aufwarten, diese Ergebnisse direkt an den Arbeitsplatz des Suchenden liefern und weil sie, und das ist besonders hervorzuheben, eine wesentlich größere Menge an Dokumenten durchsuchen, als dies mit jeder bisher bekannten Technologie möglich war. Man sollte diese Aussagen nicht falsch verstehen. Die Qualität von Suchmaschinen ist nicht gleichbedeutend mit der Leistungsfähigkeit des gegenwärtigen „Marktführers“ Google, sondern gemeint ist die Technologie Suchmaschine. Und man sollte auch nicht unterstellen, dass der leider häufig praktizierte Weg der zu kritisierenden oberflächlichen Nutzung von Suchmaschinen positiv herausgestellt wird, sondern es geht darum, dass durch den zielgerichteten Einsatz der Suchmaschinenteknologie das mögliche Retrieval deutlich bessere Ergebnisse bringt, als es die Wissenschaft bisher gewohnt war. Natürlich haben heute eingesetzte Suchmaschinen noch genügend viele ungelöste Probleme, die sie nicht annähernd an die menschliche Leistungsfähigkeit von Retrievalspezialisten in den wissenschaftlichen Bibliotheken oder Fachinformationszentren heranreichen lässt. Diese Spezialisten sind selten und ihr Einsatz ist häufig nicht wirtschaftlich. Die gegenwärtigen Forschungen auf den Gebieten Semantic Web, Information Retrieval oder Wissensmanagement lassen jedoch berechtigte Hoffnungen auf immer leistungsfähigere Suchtechnologien aufkommen.

Viel zu selten wird der mögliche Einfluss der Autorinnen und Autoren im Sinne der wahrnehmbaren Verantwortung für das Erschließen und spätere Auffinden der wissenschaftlichen Arbeit diskutiert. Im bisherigen papierbasierten System begrenzte sich ihre Rolle höchstens auf das Verfassen eines Abstracts und/oder erklärender Schlüsselwörter. In der elektronischen Welt könnten sie zusätzlich durch einen gut strukturierten Text sehr viel zur Verbesserung der Treffergenauigkeit beim Retrieval beitragen. Betrachten wir uns heute einen Text, z. B. in einer wissenschaftlichen Zeitschrift, so erkennt der Mensch ihn durchaus als wohl strukturiert, indem sich der Titel, der Name des Verfassers, die Überschriften der Kapitel und z. B. ein Zitat deutlich vom übrigen Text abheben. Diese Dokumentstruktur könnte man in der digitalen Ausprägung eines Textes auch maschinenlesbar machen, indem jede der Passagen, die sich vom „Body“ der Arbeit abhebt, gesondert kennzeichnet wird. Nichts anderes passiert durch die Nutzung der Hypertext Markup Language (HTML) bei der Darstellung von Web-Seiten. Erst durch diese Kennzeichnung (das Taggen) sind die Browser in der Lage, eine adäquate Darstellung auf den Bildschirm zu projizieren. Durch die internationale Standardisierung von HTML wurde eine Beschreibung des Dokumenttyps „Web-Präsentation“ vorgenommen. In der Zwischenzeit lässt sich die Struktur von jedem geistigen Werk digital darstellen. Auf der Basis der eingangs schon erwähnten Standard Generalised Markup Language (SGML) ist eine Vielzahl weiterer Markup-Sprachen entstanden, die jeweils spezielle Dokumenttypen beschreiben. Eine Ausnahme bildet die Extensible Markup Language (XML), mit der es ebenso wie mit SGML möglich ist, neue Dokumenttypdefinitionen (DTD) zu erstellen. Würden z. B. alle Promovenden die Dokumenttypdefinition für Dissertationen ([http://edoc.hu-berlin.de/e\\_autoren/download/dissonline.dot](http://edoc.hu-berlin.de/e_autoren/download/dissonline.dot)), die an der Humboldt-Universität entwickelt wurde, benutzen, so wäre eine qualifizierte Suche nicht nur nach den bisher gängigen Kategorien „Autor“, „Titel“ und „Schlagwörter“ möglich, sondern auch nach Kapitelüberschriften, Abbildungen,

Zitaten, Literaturverweisen u. Ä. Auf diese Weise könnte eine deutlich größere Treffergenauigkeit als bisher erreicht werden, weil die Suche spezifischer formuliert werden könnte. Detaillierter wird diese Herangehensweise in diesem Heft in den Beiträgen zu den Autoren hinweisen dargestellt.

## Neue Möglichkeiten durch das elektronische Publizieren

Neben der Erhöhung der Effizienz und der Verbesserung vorhandener Charakteristika des wissenschaftlichen Publizierens ergeben sich durch das elektronische Publizieren neue - bisher nahezu undenkbare - Gestaltungsspielräume. An erster Stelle ist sicherlich die Möglichkeit zu nennen, nicht nur textorientierte Arbeiten, die mit Grafiken illustriert sind, zu verfassen, sondern multimediale Werke zu erstellen, d. h. neben den genannten Formaten können in diese Veröffentlichung auch Audio- und Videosequenzen integriert werden. Damit eröffnet sich eine bisher nicht darstellbare Vielfalt für die wissenschaftliche Publikation. Noch gibt es nur wenige Arbeiten, vorrangig in den Kulturwissenschaften und vereinzelt in den Naturwissenschaften mit den Chemikern als Vorreitern, die diese Formen ausnutzen. Das Neue an der gegenwärtigen Situation ist nicht die Möglichkeit des Darstellens an sich, das konnte man in der „analogen Zeit“ auch über Tonbänder oder Filme, sondern die Integration in ein wissenschaftliches Werk, in dem sämtliche Bestandteile der Arbeit digital vorliegen, in einer abgrenzbaren und abgeschlossenen Einheit erfassbar sind und als solche durch Methoden der Nutzung der digitalen Signatur in ihrer Integrität und Authentizität nachweisbar.

Ein weiterer Vorteil einer elektronisch vorliegenden Publikation ist die Chance, ergänzende Informationen zu diesem Werk dem Dokument unmittelbar zuzuordnen. An erster Stelle sind hier die deskriptiven Metadaten aufzuführen, die die Suche erheblich verbessern. Zu nennen wären technische Daten zum Format, die Größe und der Zeitpunkt des Abspeicherns. Möglich ist ebenso die Angabe der jeweiligen Lizenz, unter die das Werk für eine potentielle

Nutzung durch Dritte gestellt wurde (siehe hierzu in diesem Heft den Beitrag zu Creative Commons von Manuel Klatt-Kafemann) oder ergänzende Angaben zur Langfristarchivierung. Weitere Metadaten sind denkbar, aber wenn nicht einheitlich angewandt, nicht immer unbedingt nutzbringend. Während es sich hierbei um so genannte statische Metadaten (sich in der „Lebenszeit“ eines Dokuments nicht verändernde Daten) handelt, eröffnet das elektronische Publizieren auch die Chance, dynamische Metadaten anzugeben. Gemeint sind hier Daten, die sich im Verlauf der Nutzung des Dokumentes verändern. Metadaten könnten die Downloadzahlen sein, die eine erste Einschätzung für den Lesenden über die Häufigkeit der Nutzung des Werkes eröffnet. Natürlich verändern sich diese Zahlen mit jeder Nutzung und bringen so eine Zusatzinformation (siehe hierzu in diesem Heft den Beitrag zu „Impact von Open-Access-Publikationen“ von Sabine Henneberger). Dynamische Daten über das Dokument könnten aber auch Einschätzungen des Inhaltes der Arbeit durch Lesende sein, die als Ergänzung bzw. als Kommentar und Zusatzinformation für Dritte gedacht sind. Diese und weitere Technologien, die häufig als Web 2.0 Technologien zusammengefasst werden, sind in ihrem Nutzen umstritten, aber sie sind prinzipiell möglich und können bei zielgerichtetem Einsatz sicher auch entsprechenden Mehrwert bringen.

Neben dem Gewinn, den man durch den Einsatz von Metadaten erreichen kann, sind insbesondere die „Virtuellen Verknüpfungen“ ein großer Vorteil des elektronischen Publizierens. Anders als bei der traditionellen Publikationsart, bei der ausschließlich die Möglichkeit des Zitierens zur Verknüpfung von Texten besteht, eröffnen sich die Chancen der Verlinkung von korrespondierenden wissenschaftlichen Arbeiten. Das sind zum einen ein inhaltlicher Mehrwert und zum anderen ein zeitliches Plus, da man direkt auf den gerade zitierten Beitrag verwiesen wird bzw. auf ergänzende Beiträge den entsprechenden Hinweis erhält.

## Gegenwärtige Grenzen des elektronischen Publizierens

Der Prozess des wissenschaftlichen Publizierens hat eine lange, über Jahrhunderte reichende Tradition. Da ist es gut nachvollziehbar, dass das elektronische Publizieren mit einer Entwicklungsgeschichte von etwa 18 Jahren an vielen Stellen auf Probleme stößt, für die es bisher keine befriedigende Lösung gibt. Ein Grundproblem, an dem man jedoch nicht jede Unzulänglichkeit festmachen kann, ist der nahezu durchgängige Versuch, die bisherigen Regularien beim Publizieren durch das elektronische Publizieren abzubilden. Das bedeutet letztlich, dass das auf der analogen Technologie beruhende traditionelle papierbasierte Veröffentlichen in der „digitalen Welt“ nachempfunden wird. Hervorstechendes Beispiel ist die nach wie vor vorherrschende Gepflogenheit, auch elektronische Publikationen mit entsprechenden Seitennummern, ausgerichtet z. B. am DIN A4-Format des Papiers, zu versehen. Das vorrangige Wiedergabemedium, der Computerbildschirm, ist für dieses Format nur begrenzt geeignet. Er ist vielmehr geprägt durch das 3:4 bzw. 6:9 Verhältnis der Bildschirmausmaße, bei dem sich Webseiten deutlich besser darstellen lassen. Das Festhalten an Papierformaten und der einhergehenden Seitennummerierung ist in erster Linie einer fehlenden Alternative für das eindeutige Zitieren, das beim wissenschaftlichen Publizieren von besonderer Brisanz ist, geschuldet. Dieses Problem fasst Uwe Müller (in [4], Seite 44) unter dem Begriff der Nachhaltigkeit zusammen. „Die dauerhafte Haltbarkeit einer Publikation als Voraussetzung für die nachhaltige Verfügbarkeit, Rezipierbarkeit und Referenzierbarkeit steht im Falle elektronischer Publikationen gleich in mehrfacher Hinsicht zur Disposition“. Bei der kurzen Geschichte elektronischer Medien ist es verständlich, dass zu geringe Erfahrungen vorliegen, um sagen zu können, welchen der gegenwärtig eingesetzten Speichermedien welche Lebensdauer vorhergesagt wird. Die kurz nach ihrer Einführung als die Lösung des Problems der Langfristarchivierung bezeichneten CD-ROMs sind in der Zwischenzeit diesen Ruf wieder los-

geworden, nachdem auch hier einige Defizite festgestellt werden mussten. Aber nicht die Speichermedien allein sind die Herausforderungen bei der Langfristverfügbarkeit digitaler Dokumente, sondern die Beständigkeit der Hardware, wobei hier die den technologischen Fortschritt begleitenden Fragestellungen gemeint sind, als auch die zum Einsatz kommende Software, die sich ebenso rasant entwickelt und in vielen Fällen auf die Vergangenheit keine „Rücksicht“ nimmt, im Sinne der Wahrung der Kompatibilität der seinerzeit erstellten Dokumente. Vor gut 25 Jahren unter Nutzung des damaligen Weltmarktführers unter den Textverarbeitungssystemen „Wordstar“ erzeugte Dateien sind heute nur noch mit großen Schwierigkeiten und erhöhtem Aufwand lesbar. Eine Zusammenfassung der gegenwärtigen Anstrengungen auf dem Gebiet der Langfristarchivierung elektronischer Dokumente soll hier nicht vorgenommen, sondern auf den entsprechenden Artikel von Susanne Dobratz [9] verwiesen werden.

In der nicht immer sachlich geführten Diskussion (z. B. in [10]) wird dem elektronischen Publizieren sehr häufig ein Qualitätsproblem unterstellt. Dadurch, dass jeder, der auch nur minimale technische Kenntnisse hat, etwas ins Netz stellen kann und somit eine breite Öffentlichkeit erreicht, leidet die inhaltliche Qualität.

Diesem Argument kann man zunächst nur wenig entgegenstellen. In gewissem Sinne ist dieses „ins Netz stellen“ vergleichbar mit der so genannten „grauen Literatur“, also in Papierform erscheinenden Veröffentlichungen, für die es keinen geregelten Vertriebskanal – zum Beispiel über den Buch- und Zeitschriftenhandel – gibt. Der entscheidende Unterschied ergibt sich durch den finanziellen und personellen Aufwand zur Produktion von Forschungsberichten, Studien von Fachgesellschaften, Empfehlungen von wissenschaftlich ausgerichteten Vereinen, den klassischen Beispielen für „graue Literatur“. Da für das elektronische Pendant deutlich weniger Aufwand (vor allem finanzieller) notwendig ist, sind die „Hemmschwellen“ für solche Veröffentlichungen deutlich geringer. Andererseits sollte gerade

aus den folgenden Beiträgen des cms-journals (von Gradmann, St., Weishaupt, K. und Müller, U. /Schirmbacher, P.) ersichtlich werden, dass es vielfältige Anstrengungen gibt, diesem Nachteil etwas entgegenzusetzen, wie das geregelte Peer Review bei Open-Access-Journalen, die Veröffentlichungspolicy von Repositories und das Public-Review-Verfahren als häufig unterschätzte Meinungsäußerung der kritischen Öffentlichkeit.

Die größte Unsicherheit wird jedoch durch Unklarheiten bei den Möglichkeiten der Wahrung der Authentizität und der Integrität digitaler Dokumente erzeugt. Unter der Authentizität versteht man in diesem Zusammenhang die eindeutige Zuordnung der Verfasser als Schöpfer des jeweiligen Dokuments. Im Unterschied dazu bringt die Integrität die Unversehrtheit des Dokuments seit seiner Erstellung zum Ausdruck.

Aus den Negativverfahren des Umgangs im Internet schöpfend, wird die Möglichkeit bezweifelt, tatsächlich den Urheber des Werkes zweifelsfrei benennen zu können und vor allem ein garantiert seit seiner Entstehung unverändertes „Originaldokument“ vorliegen zu haben. Man könnte meinen, dass es diese Probleme im traditionellen Publikationsverfahren nicht gibt. Das ist nicht richtig, wir alle haben vielfach nur verdrängt, dass im wissenschaftlichen Publikationskreislauf eine Instanz zwischen-geschaltet ist. Wir betrachten letztlich und sicher zu Recht, weil sie um ihren Ruf bedacht sein müssen, die Verlage als vertrauenswürdige Institutionen, die dem Lesenden garantieren, dass sowohl die angegebenen Autoren die tatsächlichen Urheber des Werkes sind und dass dieses Werk seit der Übergabe des Manuskripts an den Verlag nicht verändert wurde. Ein ähnliches Verfahren existiert in der Zwischenzeit auch für die digitale Welt. Zum einen gibt es die digitale Signatur, ausgestellt von einem staatlich anerkannten Trustcenter, die sowohl die Authentizität des Signaturinhabers belegt, als auch die Authentizität des Betreibers – zum Beispiel eines Repositories – zum Ausdruck bringt. Zur Sicherung der Integrität kennen wir zum anderen die Bildung des Hashwertes von Dokumenten auf der Basis von vertrauenswürdigen

Algorithmen, der gemeinsam mit dem Dokument abgelegt und signiert wird und somit die Unversehrtheit eines Dokumentes belegen kann. (Nähere Erläuterungen dazu sind im Artikel von Niels Fromm „Signatur und Zeitstempel zur Wahrung von Authentizität und Integrität“ in diesem Heft zu finden.)

Es ließen sich weitere Problemkreise finden, die gewisse Unzulänglichkeiten beim elektronischen Publizieren belegen könnten, wie die häufige Angabe lediglich eines Uniform Resource Locators (URL) als Hinweis auf die Position der Publikation im Netz. Nun wissen wir, dass sich die Notwendigkeit, eine solche URL zu ändern, immer mal ergibt, z. B. durch den Wechsel des Server o. Ä. und deshalb orientiert das DINI-Zertifikat auf den Einsatz von persistenten Identifiern. Dafür gibt es mehrere Verfahren, die hier jedoch nicht im Detail dargestellt werden sollen.

Anderes Beispiel ist die Zitierbarkeit von multimedialen Bestandteilen einer Publikation bzw. von Forschungsdaten, die einer Arbeit zugrunde liegen, weil es an entsprechenden Standards fehlt, die weltweit für eine Vereinheitlichung sorgen würden.

## Fazit

Es würde den Umfang eines solchen Artikels sprengen, wenn er sich zum Ziel gesetzt hätte, auf jedes Detail des Publikationsprozesses und seiner Widerspiegelung in der „digitalen Welt“ einzugehen. Es sollte zum Ausdruck kommen, dass das elektronische Publizieren erst am Anfang seiner Entwicklung steht und man somit noch mit einer Vielzahl von Unzulänglichkeiten rechnen muss, andererseits jedoch das Potential vorhanden ist, den Publikationsprozess effektiver zu gestalten und an einigen Stellen in neue technische Qualitätsdimensionen vorzustoßen. Zum gegenwärtigen Zeitpunkt muss es darum gehen, die Probleme zu erkennen, zum Gegenstand der Forschung zu machen und möglichst rasch zu gemeinsamen Standards und internationalen Regelungen zu kommen.

## Literatur

- [1] DINI-Zertifikat Dokumenten- und Publikationsservice 2007. URL: [http://www.dini.de/fileadmin/docs/dini\\_zertifikat\\_2007\\_v2.1.pdf](http://www.dini.de/fileadmin/docs/dini_zertifikat_2007_v2.1.pdf)
- [2] MÜLLER, U.; SCHIRMBACHER, P.: *Das wissenschaftliche Publizieren – Stand und Perspektiven*. cms-journal 32, Berlin 2009, S.7 – 12
- [3] SUBER, P.: *Open Access Timeline*. 2009, URL: <http://www.earlham.edu/~peters/fos/timeline.htm>
- [4] MÜLLER, U.: *Peer-Review-Verfahren zur Qualitätssicherung von Open-Access-Zeitschriften: Systematische Klassifikation und empirische Untersuchung*. Dissertation, Humboldt-Universität zu Berlin, 2008, URL: <http://edoc.hu-berlin.de/dissertationen/muelleruwe.thomas-2008-12-17/>
- [5] RIEHM, U.; BÖHLE, K.; WINGERT, B.: *Elektronisches Publizieren*. In Kühlen, S.; Seeger, T.; Strauch, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. München: K.G. Saur, 2004, S. 549–558
- [6] GRADMANN, ST.: *Verbreitung vs. Verwertung – Anmerkungen zu Open Access, zum Warencharakter wissenschaftlicher Informationen und zur Zukunft des elektronischen Publizierens*. URL: <http://edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=28934>
- [7] URL: <http://edoc.hu-berlin.de/browsing/cms-journal>
- [8] DOBRATZ, S.; MÜLLER, U.: *Wie entsteht ein Institutional Repository? – Eine systematische Hinführung in acht Schritten*. cms-journal 32, Berlin 2009, S. 47 – 54
- [9] DOBRATZ, S.: *Grundfragen der digitalen Langzeitarchivierung für den edoc-Server*. cms-journal 32, Berlin 2009, S. 93 – 98
- [10] REUSS, R.; IN F.A.Z.: „Eine heimliche technokratische Machtergreifung“ 11.02.2009; URL: <http://www.faz.net/s/RubC3FFBF288EDC421F93E22EFA74003C4D/Doc-E8A8C72C667984805A1F2CF2954CD7C78-ATpl-Ecommon-Scontent.html>

# Publizieren im Open-Access-Modell

## Allgemeine Einführung und Grundaussagen

Stefan Gradmann | stefan.gradmann@ibi.hu-berlin.de

### Was ist „Open Access“?

Die Grundidee von Open Access ist einfach: Ziel ist der freie und ungehinderte Zugang zum publizierten Wissen – hier in erster Linie dem wissenschaftlich kodifizierten Wissen. Nicht ganz so einfach vermittelbar sind die unterschiedlichen Ansätze, mit denen dieses Ziel konkretisiert und erreicht werden soll. Und recht kompliziert wird es sogar, wenn man sich Klarheit zu verschaffen versucht darüber, was sich eigentlich inhaltlich und technisch hinter den beiden Schlüsselbegriffen „Open“ und „Access“ verbirgt.

Der Begriff ‚Open‘ ist dabei schwer zu fassen und kaum zu übersetzen: Ganz offenkundig ist die wörtliche Übersetzung „offen“ irreführend. Aber auch die Übersetzung mit „frei“ löst das Problem nicht wirklich: Ist hier „kostenfrei“ gemeint – und wenn ja, wem entstehen dann keine Kosten mehr und wem sehr wohl?


Auch das Wort „access“ entpuppt sich bei näherer Betrachtung als vielschichtig: Ist allein der freie lesende Zugriff auf Bildschirminhalte schon ausreichend oder müssen gewisse technische Nachnutzungsmöglichkeiten bestehen (Ausdrucken, Weiterverwendung von Quelltext) bzw. vielleicht sogar bestimmte technische Anforderungen an die Publikationsform erfüllt sein (offene vs. proprietäre Dokumentformate)?

Fragen dieser Art ergeben zusammen mit den dargestellten unterschiedlichen Strategien der Umsetzung von Open Access eine recht komplexe Matrix von Fragen und Problemlagen, die eine schnelle und flächendeckende praktische Ausbreitung des Open-Access-Gedankens bislang jedenfalls nicht gerade gefördert hat!

### Zur Geschichte von Open Access

Seinen Ursprung hat der Open-Access-Gedanke in gewisser Hinsicht in der Physik, und dort spezifischer bei Paul Ginsparg vom Los Alamos National Laboratory, der erstmals 1991 im großen Maßstab die Idee umsetzte, für die Verbreitung von Preprints wissenschaftlicher Aufsätze einen eigens zu diesem Zweck aufgesetzten Server, das sogenannte ArXiv verfügbar zu machen und zugleich potentiellen Beiträgern über eine komfortable Softwarelösung das Einstellen und Suchen solcher zentral archivierter Preprints zu ermöglichen.

Ginsparks Idee war eine veritable Innovation, die rein technisch gesehen eines der enormen schlummernden Potentiale der Internettechnologie nutzte, und diese Idee fand in der Folge in unterschiedlichen Disziplinen Nachahmer aufgrund ihrer schieren Effizienz. Der nun folgende Aufbau weiterer fachbezogener und institutioneller Server zusammen mit der zunehmenden Verbesserung der dazu verfügbaren Softwarelösungen sowie der Vereinheitlichung von Lösungen für das „Harvesting“ von Publikations-Metadaten führten zu Angeboten wie E-Prints (als generische Software-Lösung aus dem Angebot des Electronic and Computer Science Department der University of Southampton) und zur Entstehung der Open Archives Initiative, die mit der ersten Version ihres Protokolls für das „Metadata Harvesting“ (PMH 1.0) im Jahr 2000 einen ersten Ansatz für die Interoperabilität dieser verteilten Serveransätze präsentierte.

 Beitrag gibt, ausgehend von einem ersten Definitionsversuch, einen Abriss zur Geschichte von „Open Access“. Er beschreibt die beiden Hauptvarianten der Open-Access-Bewegung (OA „Grün“ vs. OA „Gold“) und benennt die Hauptproblempunkte und aktuellen Diskussionsbereiche im Open-Access-Publizieren. Der Beitrag schließt mit der Benennung derjenigen Felder in der wissenschaftlichen Welt, in denen für die dauerhafte Durchsetzung des Open-Access-Gedankens vorrangig Handlungsbedarf gegeben ist.

OA „Grün“, OA „Gold“, wissenschaftliche Reputation, Publikationsformate, Berliner Erklärung, Budapest Declaration, Bethesda Statement



Dennoch waren diese Ansätze noch weit entfernt von der wenige Jahre später einsetzenden enormen medialen Präsenz von „Open Access“, das als Begriff aus dem wissenschaftlichen Publikationswesen eine Öffentlichkeitswirksamkeit entfaltet hat wie kaum ein Begriff aus diesem Umfeld zuvor.

Ursache dafür war eine Konjunktur von Strömungen: Die in den 90ern aufgebaute Infrastruktur von Dokumentenservern traf um die Jahrtausendwende zusammen mit einer tiefen Unzufriedenheit vor allem der Bibliothekare mit dem, was damals als die „Serials Crisis“ apostrophiert wurde: das zunehmende Auseinanderklaffen der Preisentwicklung wissenschaftlicher Zeitschriften (fast ausschließlich im Bereich der „harten“ Naturwissenschaften) einerseits und der begrenzten Leistungsfähigkeit der wissenschaftlichen Bibliotheksetats andererseits, was angesichts der Monopolstellung einzelner Großverleger wie Elsevier für die Bibliotheken zum existentiellen Problem wurde.

Hinzu kam, dass die traditionellen „geschlossenen“ Publikationskanäle der etablierten Wissenschaftsverleger mehrheitlich nicht eingerichtet waren auf die Möglichkeiten, die sich aus den vernetzten, verteilten Informationsstrukturen des WWW für das elektronische Publizieren ergeben, sofern alle Akteure ein Mindestmaß an technischer und inhaltlicher Offenheit an den Tag legen: Die geschlossenen Publikationsketten großer Anbieter wurden nun zugleich als monopolistisch, unbezahlbar und technisch ineffektiv empfunden – und diese Wahrnehmung zusammen mit der in den 90er Jahren aufgebauten ersten Generation einer alternativen technischen Infrastruktur führte als Syndrom zur Entstehung der Open-Access-Bewegung.

Es entstanden nun sehr bald Institutionen übergreifende Initiativen wie SPARC und SPARC-Europe, innovative und z. T. auch kommerziell erfolgreiche Publikationsplattformen wie die Public Library of Science (PLOS) oder Biomed Central (BMC).

Vor allem entstand nun in rascher Folge eine Vielzahl stark wissenschaftspolitisch motivierter Resolutionen und Erklärungen (zumindest in dieser Hin-

sicht war die Open-Access-Bewegung aus dem Stand unglaublich produktiv!). Die zumindest in der kanonischen Sichtweise erste dieser Erklärungen verdankt sich einer durch das OSI (Open Society Institute), im Dezember 2001 in Budapest organisierten Konferenz, deren Ziel die Zusammenführung der bereits existierenden unterschiedlichen Open-Access-Aktivitäten war, zusammen mit dem Versuch, zu präzisieren, für welche Arten wissenschaftlicher Literatur der freie Zugang nun eigentlich möglich gemacht werden sollte.

Der daraus resultierende Initiativaufruf erschien am 14.02.2002 und fasst die Ziele von Open Access folgendermaßen zusammen: „An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.“

Der Aufruf bleibt also klar beschränkt auf Zeitschriftenaufsätze und thematisiert auch noch mit keinem Wort die publikationsökonomische Seite von Open Access: Es geht wirklich um die maximale Verfügbarkeit und Sichtbarkeit von Zeitschriftenbeiträgen, nicht mehr und nicht weniger.

Das sollte sich mit dem folgenden Großtreffen der Gemeinde und dem daraus resultierenden Bethesda Statement on Open Access Publishing, das am 20. Juni 2003 veröffentlicht wurde, grundlegend ändern, denn hier lag nunmehr der Schwerpunkt ganz eindeutig im Bereich des elektronischen Publizierens im Open-Access-Modell. In diesem Dokument findet sich eine erste Definition einer Open-Access-Publikation:

„An Open Access Publication is one that meets the following two conditions:

1. The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly

and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.

2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.“

Das Bethesda Statement thematisiert also das elektronische Publizieren im Open Access Modell und enthält – zumindest explizit – auch keine Einschränkung mehr auf Zeitschriftenaufsätze.

Die zeitlich unmittelbar darauf folgende sog. „Berliner Erklärung“ nun stellt Open Access definitiv in einen großen, wissenschaftspolitischen Zusammenhang und war das Ergebnis einer von der Max-Planck-Gesellschaft organisierten Konferenz deutscher und internationaler Wissenschafts- und Förderorganisationen. Die „Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“ spannt einen großen Bogen über alle Spielarten von Open Access und alle Wissenschaftskulturen in der folgenden, sehr allgemeinen Fassung des Begriffs „Open Access“ und der damit verbundenen Forderung, dies Paradigma zu unterstützen: „New possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported. We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community.“

Die Berliner Erklärung bleibt in der Folge hinsichtlich dessen, was nun eigentlich offen zugänglich gemacht werden soll, zwar weniger deutlich als die beiden Vorgängererklärungen und operiert stattdessen mit dem sehr weit gefassten Begriff der „Open Access Contribution“, doch

sind die dann auf der Folgekonferenz 2005 in Southampton beschlossenen Umsetzungsempfehlungen schon klarer: „In order to implement the Berlin Declaration institutions should:

1. implement a policy to require their researchers to deposit a copy of all their published articles in an open access repository and
2. encourage their researchers to publish their research articles in open access journals where a suitable journal exists and provide the support to enable that to happen.“

Die Berliner Erklärung hat Open Access zum wissenschaftspolitischen Großthema gemacht, und die seitdem jährlich stattfindenden Folgekonferenzen betrafen dann unterschiedliche Detailaspekte des mit „Berlin“ endgültig etablierten Großthemas wie etwa eine Implementation Roadmap als Ergebnis von Berlin2 (2004) an der Großforschungseinrichtung CERN, die schon zitierte Umsetzungsempfehlung als Ergebnis von Berlin3 (2005).

## „Green“ vs. „Gold“: unterschiedliche Wege zu Open Access

Der oben gegebene historische Abriss sollte deutlich gemacht haben, dass die Open-Access-Bewegung bei allem Erfolg und aller nunmehr auch politischer Akzeptanz lange Jahre an der Dichotomie zweier Ausprägungen des Grundgedankens laboriert hat, die plakativ immer wieder als der „goldene“ und der „grüne“ Weg zu Open Access bezeichnet wurden.

Dabei meint Open Access „green“ die parallele Veröffentlichung anderweitig bereits publizierter Zeitschriftenaufsätze auf persönlichen, institutionellen oder fachbezogenen Servern und Repositorien. Unter Open Access „gold“ hingegen wird die primäre Publikation eines wissenschaftlichen Dokuments in einer Open-Access-Zeitschrift verstanden.

Darüber, inwieweit beide Ansätze als konkurrierend oder als komplementär zu sehen sind, hat es vor allem in den ersten Jahren der Bewegung zum Teil

heftige Polemiken gegeben, insbesondere zwischen den beiden exponierten Protagonisten Stevan Harnad (grün) und Jean Claude Guéron (gold), die jeweils eine der beiden Linien für allein seligmachend zu erklären versuchten.

Doch ist es um diese Kontroverse deutlich ruhiger geworden, und allgemein scheint sich eine friedliche Koexistenz der beiden Ansätze etabliert zu haben, die eher pragmatisch Komplementarität dort ausreizt, wo sie deutlich erkennbar ist und latente Zielkonflikte (die es durchaus gibt) nicht mehr offen und polemisch diskutiert.

An die Stelle der zuletzt stellenweise schon eher sterilen Diskussion um ‚green‘ vs. ‚gold‘ sind nun eine Reihe neuer Problempunkte und Diskussionsbereiche getreten, die derzeit in der Open-Access-Bewegung diskutiert werden, und um die es im folgenden Abschnitt gehen soll. Dabei beschränke ich mich nunmehr ausdrücklich auf Fragen des elektronischen Publizierens im Open-Access-Modell und kann die teilweise sehr komplexen Fragestellungen an dieser Stelle auch nur grob andeuten.

## Hauptproblempunkte und aktuelle Diskussionsbereiche im Open-Access-Publizieren

Open Access war bislang ganz ausschließlich auf das Publikationsformat Zeitschrift und die darin enthaltenen Aufsätze fokussiert. Dies schließt weite Teile vor allem der Geistes- und Sozialwissenschaften aus, die nicht so stark wie etwa die Physik durch das Publikationsformat Zeitschrift bestimmt sind, und deren eher monographische Veröffentlichungsformate ganz generell aufgrund ihrer inhärenten Komplexität andere technische und ökonomische Anforderungen an elektronische Publikationsmodelle stellen.

Hierbei stellen sich spezifisch auf Open Access bezogen insbesondere Fragen im Zusammenhang mit dem zweiten Diskussionsbereich der Publikationsökonomie.

Veröffentlichungen im Open-Access-Modell kosten Geld. Weniger vielleicht als kommerzielle Publikationswege – aber

einer Refinanzierung bedürfen sie allemal. Die Frage, welche Geschäftsmodelle für eine Open-Access-basierte Publikationsökonomie eigentlich angemessen und überlebensfähig sind, spielte schon früh eine gewichtige Rolle in der Diskussion, wobei es lange Jahre vor allem um die Umleitung der Kosten für Zeitschriftenaufsätze vom Abnehmer (v. a. Bibliotheken) zum Verursacher der Publikation ging (Stichwort ‚author pays‘).

Diese an sich schon schwierige Diskussion ist durch das Hinzukommen weiterer Publikationsformate (s. o.) noch deutlich komplexer geworden.

Eine inzwischen sehr angeregte Diskussion wird generell um die Frage geführt, inwiefern bestimmte digitale wissenschaftliche Publikationen überhaupt noch rezipierbar und überprüfbar sind, ohne dass die ihnen zugrunde liegenden Primärdaten (Messwerte etc.) mit publiziert werden. Eine radikalere Form dieser Fragestellung geht dahin, ob auch die Veröffentlichung strukturierter Datensammlungen selbst als eine wissenschaftliche Publikation sui generis anzusehen ist.

Beide Aspekte der Fragestellung haben im Kontext von Open Access insofern eine besondere Note, als Wissenschaftler das Prinzip auch der Primärdatenpublikation zwar grundsätzlich begrüßen, im konkreten Fall jedoch aus unterschiedlichen Gründen sehr zurückhaltend reagieren, wenn es um die Umsetzung geht – sollen doch hier Informationen öffentlich zugänglich gemacht werden, die in der Vergangenheit eben nicht öffentlich waren!

Eine weitere, schon länger andauernde Diskussion dreht sich um die Frage, welche Rahmenbedingungen und Anreizsysteme eigentlich geschaffen werden müssen, um vor allem auch Nachwuchswissenschaftler zur Veröffentlichung im Open-Access-Modell zu bewegen. Ausgangspunkt ist dabei die eher enttäuschende Tatsache, dass es bislang nicht gelungen ist, die Open-Access-Publikation als Standard-Veröffentlichungsmodell zu etablieren. Dies hat – zumindest im Fall der Geisteswissenschaften – sicher auch mit dem allgemein noch nicht wirklich etablierten Status der elektronischen Publikation zu tun. Doch gibt es auch Open-Access-spezifische Aspekte, die zu dem



Schluss führen, dass erst mit der Schaffung neuartiger Anreizsysteme eine massive Akzeptanz für das Publizieren im Open-Access-Modell erwartet werden kann.

Diese Frage der Anreizsysteme nun ist ihrerseits eng verbunden mit dem im nächsten Absatz anzusprechenden Zusammenhang zwischen Veröffentlichung und wissenschaftlicher Reputation.

Wissenschaftliche Publikation ist das zentrale Mittel für den Aufbau eigener Reputation eines Wissenschaftlers und seiner Heimatinstitution. Nun sind aber die Instrumente, mit denen die Relevanz wissenschaftlicher Publikation evaluiert wird, nicht einmal auf elektronische Publikationsformen wirklich eingerichtet, geschweige denn auf das Publizieren im Open-Access-Modell. Schlimmer noch: die wichtigsten dieser Instrumente stehen entweder den kommerziellen Verlegern zumindest nahe (wie im Falle von Thomson/ISI) oder sind sogar direkt von ihnen abhängig (wie die Elsevier gehörende wissenschaftliche Suchmaschine Scirus).

Es ist unter diesen Umständen bislang nicht gelungen, die enge Verbindung zwischen Systemen für die Reputationsbemessung und kommerziellen Publikationswegen aufzubrechen, was wohl eine weitere essentielle Voraussetzung dafür wäre, die Publikation im Open-Access-Modell ernsthaft etablieren zu können. Wie dies zu erreichen ist, stellt eines der großen Diskussionsthemen in der Open-Access-Bewegung dar!

Und schließlich wird zunehmend eine weitere Gretchenfrage diskutiert, die mit mehreren oben erwähnten Diskussionssträngen eng verbunden ist: Ist Open-Access-Publizieren möglicherweise auch in Kooperationsmodellen mit kommerziellen Verlegern vorstellbar? Diese Diskussion wäre in der Frühzeit der Bewegung sicher völlig undenkbar gewesen: viel zu ausgeprägt waren die wechselseitigen Feindbilder und viel zu ideologisch belastet die harte Konfrontation beider Lager. Doch seit zunehmend auch kommerzielle Verleger mit Open-Access-Modellen zu experimentieren beginnen und zugleich in der Open-Access-Community zusammen mit dem Nachdenken vor allem über Geschäftsmodelle eine realistischere Sicht der eigenen Möglichkeiten und Grenzen eingekehrt ist, kommt Be-

wegung in die Fronten und die Lagermentalität beginnt sich aufzulösen.

Konstellationen wie in „Publishing and the Ecology of European Research“ (PEER)<sup>1</sup> jedenfalls, wo kommerzielle Verleger und Repositoriumsbetreiber an einem gemeinsamen Projekt beteiligt sind, wären noch vor wenigen Jahren unvorstellbar gewesen!

## Handlungsbedarf in der wissenschaftlichen Welt

Der Handlungsbedarf ergibt sich unmittelbar aus den oben angesprochenen Diskussionsbereichen: Sie alle sind mit unterschiedlichen Betonungen auch Handlungsbereiche für die unterschiedlichen beteiligten Gruppen.

### Für Institutionen

Institutionen sind in erster Linie aufgefordert, das stark wissenschaftspolitisch unterfütterte System der Reputationsbemessung im Sinne von Open Access zu beeinflussen und Anreize für das Publizieren in Open-Access-Modellen zu schaffen. Ob und wie weit es Sinn macht, Open-Access-Verlage bei Einzelinstitutionen ins Leben zu rufen, ist eine offene Frage – obwohl der relative Erfolg einiger weniger universitärer Open-Access-Verlage in Deutschland dafür sprechen könnte.

### Fachgesellschaften

Auch die Fachgesellschaften sind in der Lage, die Systeme der Reputationsbemessung massiv zu beeinflussen und sollten sich zudem – wo immer denkbar – auch als Publikationsinstanzen im Open Access positionieren: Sie vor allem können die kritische Masse an Fachpublikationen anziehen, die bei rein ökonomischer Betrachtung Open-Access-Publikationsplattformen überlebensfähig machen können.

## Verlage

Verlage sollten ihre Geschäftsstrategien darauf hin überprüfen, ob nicht mit gewandelten Geschäftsmodellen auf der Grundlage von Open Access eine neue Grundlage für verlegerisches Handeln in Kooperation mit den Wissenschaftlern geschaffen werden kann – dies wird allerdings einher gehen müssen mit einer tiefer gehenden Transformation aus dem bislang praktizierten Modell der Verwertung wissenschaftlicher Publikationen in ein Dienstleistungsmodell rund um den Publikationsprozess, der selbst nicht mehr notwendig als ‚profit center‘ gesehen werden kann.

### Einzelwissenschaftler

Einzelwissenschaftler schließlich sind natürlich vor allem aufgefordert, die Vorteile einer Publikation im Open-Access-Modell wahrzunehmen und dementsprechend Publikationsentscheidungen zu treffen. Auf die Akzeptanz des Publizierens im Open-Access-Modell bei dieser Gruppe kommt alles an – doch sind sie in der Regel eher nicht in der Lage, die dafür erforderlichen Rahmenbedingungen aus eigener Kraft zu schaffen!

## Vertiefung und Praxisbeispiele

Die nachfolgenden Beiträge dieser Sektion vertiefen die hier angerissenen Themen und Aspekte des Publizierens im Open-Access-Modell bzw. geben konkrete Beispiele. So vertieft der Beitrag von Weishaupt das Thema generell unter besonderer Berücksichtigung der Geschäftsmodelle und der Qualitätssicherung. Die Beiträge von Uwe Müller gehen auf konkrete Beispiele für Open-Access-Zeitschriften ein, die auf dem eDoc-Server verfügbar sind und vertiefen Fragen der Qualitätssicherung. Der Beitrag von Hohls dann bietet einen beispielhaften Einblick in Open-Access-Publizieren in einer Geisteswissenschaft. Dobratz und Schröder schließlich gehen auf die Perspektiven eines möglichen Universitätsverlages der HU ein.

<sup>1</sup> <http://www.stm-assoc.org/press-releases/2008.10.%20PEER%20Press%20Release%20-14%20Oct%202008.pdf>

# Überblick zum goldenen Weg zu Open Access in Deutschland und international

Karin Weishaupt | weishaupt@iat.eu

## Der goldene Weg

Wer an den goldenen Weg zu Open Access denkt, hat in erster Linie Zeitschriften vor Augen, deren Beiträge im Internet frei zugänglich sind und in der Regel rein elektronisch vorliegen. Open-Access-Zeitschriften sind ein Spezialfall elektronischer Zeitschriften, der sich dadurch auszeichnet, dass für die Leser(innen) keine Kosten über die Gebühren für den Internet-Zugang hinaus anfallen und dass ihnen weitgehende Nutzungs- und Verbreitungsrechte eingeräumt werden.

Ein Teil dieser Open-Access-Zeitschriften ist gleich nach diesem Vertriebsmodell gegründet worden, der andere ist von einem ursprünglich kommerziellen Geschäftsmodell, das über Abonnementskosten finanziert wird, nach Open Access überführt worden.

Die Diskussion um Open Access hat das Bewusstsein für die Kosten wissenschaftlichen Publizierens geschärft, denn auch diese Zeitschriften erzeugen Kosten, die über verschiedene Geschäftsmodelle aufgebracht werden müssen [1]. So erfolgt die Finanzierung bei einem Teil der Open-Access-Zeitschriften durch Gebühren, die die Autor(inn)en für die Veröffentlichung ihrer Artikel zahlen. Diese können recht erheblich sein, insbesondere wenn kommerziell tätige Verlage wie Springer und Elsevier Zeitschriften anbieten, die für die Leser(innen) frei zugänglich sind. Deutlich niedriger liegen die Autorengebühren meist bei Verlagen, die sich auf Open Access spezialisiert haben, wie z. B. beim Copernicus-Verlag, der einige der renommiertesten deutschen Open-Access-Zeitschriften herausgibt. Neben diesem Author-pay-Modell kommt

die Finanzierung über Fachgesellschaften vor, wie sie z. B. in der Medizin von „German Medical Science“ praktiziert wird. Eine dritte Möglichkeit besteht in der Finanzierung durch öffentliche Geldgeber. Ein Beispiel dafür ist die Initiative „Digital Peer Publishing NRW“, die vom Land Nordrhein-Westfalen getragen wird. Die Koordination und die technische Betreuung liegen beim Hochschulbibliothekszentrum des Landes NRW, die Redaktionen der DiPP-Zeitschriften sind an den Hochschulen – oft in den Bibliotheken in Zusammenarbeit mit den Rechenzentren – angesiedelt. Schließlich stellt sich die Möglichkeit von Konsortialmodellen, wie sie zum Beispiel zurzeit in der Hochenergiephysik mit SCOAP<sup>3</sup> praktiziert wird [2].

## Informationsquellen

Wer sich lediglich über die Titel von Open-Access-Zeitschriften und nicht die darin enthaltenen Beiträge informieren will, kann dafür die elektronische Zeitschriftenbibliothek Regensburg [3] benutzen. Bei der Suche nach frei zugänglichen Zeitschriften erhält man alle, die von den teilnehmenden Bibliotheken erfasst worden sind, ohne Differenzierung nach Qualität, wissenschaftlicher oder praxisorientierter Ausrichtung oder sonstigen Merkmalen. „Ulrich's periodicals directory“ [4] erlaubt in der „advanced search“ eine Suche nach den verschiedensten Kriterien, von denen eins „Open Access“ darstellt; die Suche nach diesem Aspekt ohne weitere Einschränkung lieferte am 17.12.2008 international 3.714 Titel. Bemerkenswert ist dieses

*Ende 2008 waren 1,57 % aller Zeitschriften Open Access; bei Betrachtung der absoluten Zahlen oder des prozentualen Anteils an der gesamten Zeitschriftenproduktion stellt sich die internationale Verteilung unterschiedlich dar. Ein Problem für die Akzeptanz bilden das oft geringe Renommee und der fehlende Impact Factor. Besonders positiv wirkt die größere Publikationsgeschwindigkeit; verstärkte Anreize durch Arbeitgeber und die Betonung des kollektiven Nutzens wären weitere Fördermöglichkeiten.*

*Open-Access-Zeitschriften, goldener Weg, Akzeptanzprobleme, Autorenbefragung, Impact Factor, Fördermöglichkeiten*

Verzeichnis, weil es auch die Suche nach Zeitschriften mit Begutachtungsverfahren oder mit dem von Thomson Scientific [5] gelieferten „journal citation report“, nach Fachgebieten, Erscheinungsländern etc. und damit differenzierte Analysen erlaubt.

Die in der deutschen Open-Access-Literatur am häufigsten genannte Informationsquelle dürfte das von der Universitätsbibliothek Lund aufgebaute „Directory of open access journals – DOAJ“ [6] darstellen, das Mitte Dezember 2008 3.791 Zeitschriften verzeichnete, von denen ein gutes Drittel auch auf Artikel-ebene durchsuchbar ist.

Weniger bekannt ist in Deutschland das indische Open-Access-Zeitschriften-Portal „Open J-Gate“ [7], das am 18.12.2008 immerhin 4.710 Titel nachwies und die Recherche nach verschiedenen Kriterien erlaubt.

Noch wesentlich mehr Informationen und Suchfunktionen umfasst „redalyc“ [8], das von der Universität Mexiko aufgebaute „Red de revistas científicas de América Latina y el Caribe, España y Portugal“, das den spanischen und portugiesischen Sprachraum abdeckt und neben der Suche nach Zeitschriftentiteln eine Artikelrecherche sowie umfassende Statistikfunktionen anbietet. Damit wird zugleich die Nutzung der Zeitschriften und der einzelnen Artikel umfassend dokumentiert. Wegen der geografischen Beschränkung waren darin am 18.12.2008 zwar nur 550 Zeitschriften nachgewiesen, aber immerhin 112.135 Artikel; das DOAJ weist für die ganze Welt gerade doppelt so viele Artikel nach. Ein Portal mit einer vergleichbaren Funktionalität stellt für Europa oder konkret für Deutschland ein Desiderat dar, das die Sichtbarkeit von Open-Access-Zeitschriften und -Artikeln deutlich erhöhen und damit voraussichtlich auch deren Bekanntheit und Akzeptanz steigern könnte.

## Der aktuelle Stand

Laut DOAJ gibt es mit Stand vom 17.12.2008 in 93 Ländern Open-Access-Zeitschriften, 25 Länder haben 30 und mehr:

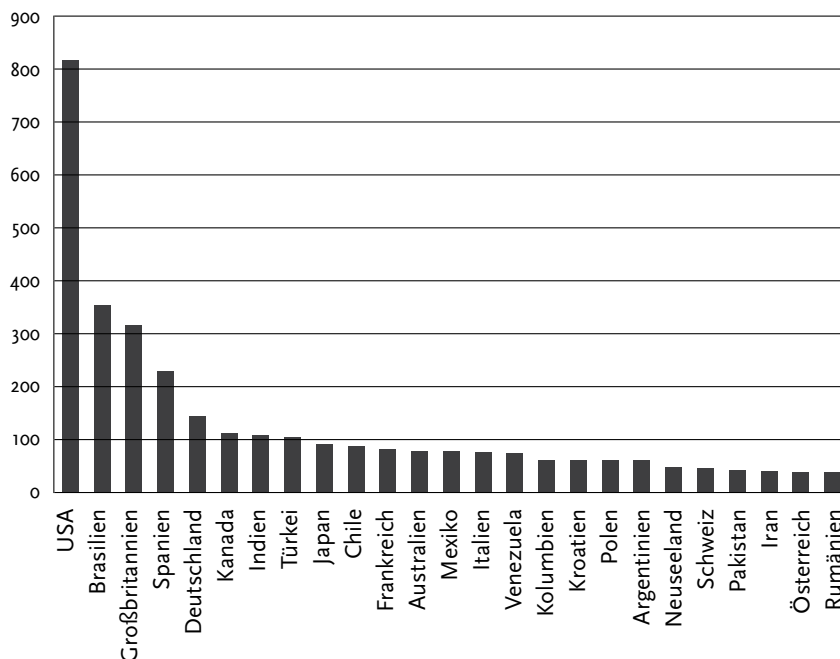


Abb. 1: Open-Access-Zeitschriften international in absoluten Zahlen; Quelle: DOAJ, 17.12.2008

Deutschland belegt in dieser Aufstellung den beachtlichen fünften Platz. Während in der Statistik der Open-Access-Repositories Deutschland mit Großbritannien um den zweiten Platz hinter den USA konkurriert, fällt hier die gute Position Brasiliens und Spaniens auf.

Um den Stellenwert der Open-Access-Zeitschriften in den einzelnen Ländern angemessen zu beurteilen, muss man zusätzlich zur Betrachtung der absoluten Zahlen diese mit der gesamten Zeitschriftenproduktion in Beziehung setzen. Dazu gibt „Ulrich's Web“ Auskunft, das sich bezüglich der Angaben für Open-Access-Zeitschriften nur geringfügig von denen des DOAJ unterscheidet. Danach hat zum Beispiel Japan etwa so viele Open-Access-Zeitschriften wie die Türkei (93 bzw. 97), die Gesamtzahl aller Zeitschriften unterscheidet sich aber um mehr als den Faktor 10 (6.230 in Japan, 610 in der Türkei). Das führt zu der Frage, welchen Stellenwert Open Access im Verhältnis zur Zahl aller wissenschaftlichen Zeitschriften und zur Gesamtzahl hat, und dabei verändert sich das Bild signifikant (siehe Abb. 2).

Ein hoher prozentualer Anteil an Open-Access-Zeitschriften ist bei niedrigen absoluten Zahlen international von nur geringer Bedeutung – erst recht, wenn deren Artikel in einer der schwach

verbreiteten Sprachen geschrieben sind –, sodass sich ein valides Urteil erst aus der Zusammenschau beider Statistiken ergibt. Da Brasilien mit 14,82 % Open-Access-Zeitschriften hinter der Türkei, die einen Anteil von 15,90 % aufzuweisen hat, wiederum den zweiten Platz belegt, ist diesem Land besondere Beachtung zu schenken.

Generell sollte nicht übersehen werden, dass es auch in den Schwellenländern bemerkenswerte Open-Access-Initiativen gibt, die für die entwickelten Länder zum Vorbild werden können. Folglich ist das Argument für Open Access, dass damit die ärmeren Länder unterstützt werden, die sich keine teuren Zeitschriftenabonnements leisten können, zwar sicher richtig, aber zu einseitig.

International machen Open-Access-Zeitschriften 1,57 % der gesamten Zeitschriftenproduktion aus. Deutschland liegt mit 0,83 % deutlich darunter.

Wenn man nicht die Gesamtzahl aller Zeitschriften als Berechnungsgrundlage wählt, sondern lediglich die wissenschaftlichen Zeitschriften, steigt der Prozentsatz der Open-Access-Zeitschriften international auf ca. 5,4 % und in Deutschland auf ca. 2 %, aber auch diese Werte stellen nicht recht zufrieden, wenn man die diversen Open-Access-Erklärungen und -Aktivitäten der letzten Jahre betrachtet.

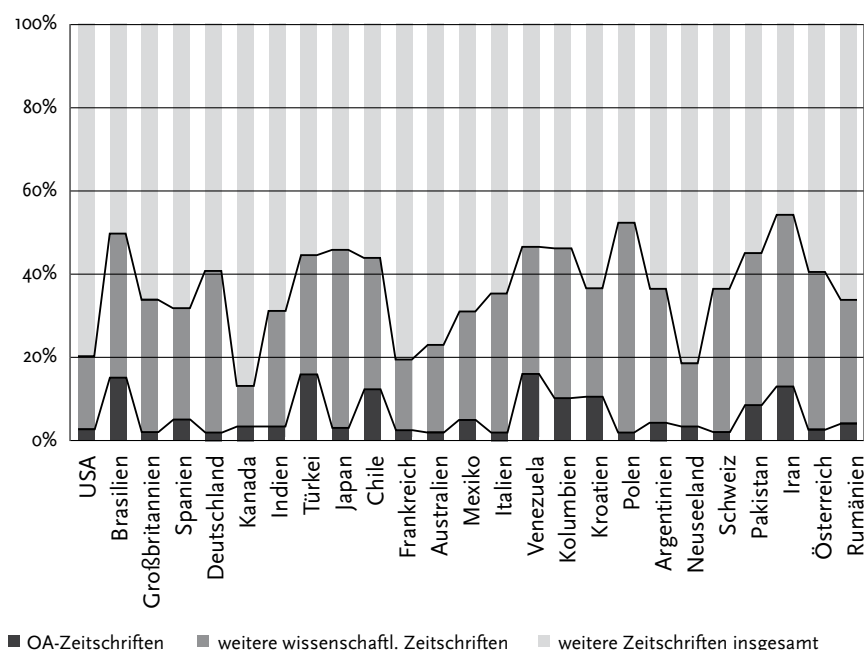


Abb. 2: Aktive Open-Access-Zeitschriften im Verhältnis zur gesamten Zeitschriftenproduktion, Quelle: Ulrich's Web, 17.12.2008

## Probleme von Open-Access-Zeitschriften

Neben der Frage, wie Open-Access-Zeitschriften zu finanzieren sind, stellt sich das Problem der Akzeptanz seitens der Autor(inn)en. Dazu ist im Mai / Juni 2008 eine Befragung unter ca. 1000 Open-Access-Autor(inn)en durchgeführt worden, aus der Maßnahmen zur Akzeptanzsteigerung abgeleitet werden sollten [9].

In der Rolle des Lesers/der Leserin hat kaum jemand etwas gegen Open Access einzuwenden, im Gegenteil: Es ist angenehm und förderlich für die wissenschaftliche Arbeit, einen benötigten Aufsatz ohne zeitliche Verzögerung und ohne Kosten am Bildschirm aufrufen zu können. Bei einem Perspektivwechsel stellt sich die Sachlage leider anders dar: Wer als Herausgeber(in) einer Zeitschrift andere als Autor(inn)en gewinnen will, wird mit einer Fülle von Vorbehalten konfrontiert: Im Falle von Autor-pay-Geschäftsmodellen sind viele nicht bereit, die geforderten Autorengebühren zu zahlen, obwohl auch bei konventionellen Veröffentlichungen oft genug die Druckkosten selbst aufgebracht werden müssen. Es werden Zweifel an der Langfristverfügbarkeit und an der Sicherung der Authentizität und Integrität der Texte

geäußert. Teilweise existieren erhebliche Informationsdefizite: Open-Access-Zeitschriften und Verfahrensfragen sind nicht ausreichend bekannt.

Häufig werden rechtliche Bedenken geäußert: Es besteht die Befürchtung, dass der eigene Text ohne Nachweis zitiert und damit missbraucht wird; dabei unterscheidet sich der Urheberrechtsschutz für gedruckte und elektronische Texte nicht, lediglich die Verwertungsrechte sind anders geregelt. Für Autorinnen und Autoren ist das Open-Access-Publizieren günstiger, da sie dabei alle Rechte an ihren Texten behalten, während die Verwertungsrechte beim konventionellen Publizieren in der Regel zumindest teilweise an die Verlage abgetreten werden; dieser Vorteil ist vielen potenziellen Autorinnen und Autoren nicht bekannt oder bewusst.

Die größte Hürde stellt die Renommeefrage dar: Da die Steigerung der eigenen wissenschaftlichen Reputation die wohl stärkste Motivation zum Publizieren darstellt, wollen die meisten Wissenschaftlerinnen und Wissenschaftler bevorzugt in renommierten Zeitschriften publizieren. Es gibt zwar keinen prinzipiellen Grund für die Annahme, dass Open-Access-Zeitschriften qualitativ schlechter sind als gedruckte oder kom-

merziell vertriebene; aber gerade Zeitschriften, die noch relativ jung sind und sich folglich in der wissenschaftlichen Welt noch nicht ausreichend etabliert haben, müssen gegen erhebliche Akzeptanzprobleme kämpfen.

Während der Ruf nach Qualitätssicherung quer durch alle Fachgebiete zu beobachten ist, konkretisiert sich in den Naturwissenschaften und der Medizin das Problem des mangelnden Renommées am Impact Factor. Dieser wird von Thomson Scientific aus der Zahl der Zitationen auf die Beiträge einer Zeitschrift ermittelt und jedes Jahr neu vergeben. Der Impact Factor ist inzwischen heftig umstritten:

- Da er sich auf die Zeitschrift insgesamt bezieht, sagt er nichts über die Rezeption und Qualität des einzelnen Artikels aus.
- Er berücksichtigt lediglich Zitate aus Zeitschriften, die ihrerseits im „Web of Science“ enthalten sind, nicht aber aus anderen Zeitschriften, Monographien, Tagungsbeiträgen, Beiträgen in sonstigen Sammelbänden und Internet-Quellen.
- Er ist leicht durch Selbstzitationen zu manipulieren.
- Da nur Zitate aus den letzten beiden Jahren berücksichtigt werden, haben es Zeitschriften aus Fachgebieten, deren Verwertungszyklen länger sind, schwerer als solche, in denen überwiegend aktuelle Literatur zitiert wird.
- Englischsprachige Journale sind überproportional vertreten, Zeitschriften in anderen Sprachen haben kaum eine Chance, einen Impact Factor zugewiesen zu bekommen.
- Spezialisierte Zeitschriften, die sich an eine kleine Community wenden, sind ebenfalls systematisch benachteiligt; Journale mit Überblicksartikeln sind klar im Vorteil. [10]

Trotzdem stellt der Impact Factor beim gegenwärtigen Stand in vielen Fachgebieten das Maß dar, an dem sich jede Zeitschrift zu messen hat.

In Brasilien gibt es mit Stand vom 18.12.2008 laut „Ulrich's Web“ 29 Zeitschriften mit Impact Factor, mit lediglich zwei Ausnahmen sind diese frei zugänglich. Damit ist dieses Land in der

glücklichen Lage, dass Impact Factor und Open Access keinen Widerspruch darstellen. In Deutschland stellt sich die Situation anders dar: Unter den 549 Zeitschriften, die einen Impact Factor aufweisen, sind nur 10 der existierenden 137 Open-Access-Zeitschriften! Kein Wunder also, dass Autor(inn)en häufig andere Zeitschriften bevorzugen und Open-Access-Zeitschriften als minderwertig ansehen! Hier gilt es, entweder um die Aufnahme der Open-Access-Zeitschriften ins „Web of Science“ und damit die Zuteilung des Impact Factors zu kämpfen oder der Fachöffentlichkeit bewusst zu machen, dass diesem Wert eine unverhältnismäßig hohe Bedeutung zugemessen wird und dass alternative Methoden des Qualitätsnachweises Vorzug bekommen sollten – ein Ziel, das nicht so schnell zu erreichen sein wird!

## Motivation für Autorinnen und Autoren

Eine Hilfe für die Umsetzung des Open-Access-Gedankens wäre es, wenn Open-Access-Veröffentlichungen bei der Evaluation von Forschungsleistungen besonders gewürdigt oder sonstige Anreize geschaffen würden – durch Arbeitgeber, Förderinstitutionen und Einrichtungen der Wissenschaftsevaluation. Während die Harvard University seit Frühjahr 2008 das Open-Access-Publizieren für verbindlich erklärt und Abweichungen explizit begründet werden müssen und die Universität Zürich das Vorhandensein einer frei zugänglichen Version für die Aufnahme eines Textes in die Hausbibliographie zur Voraussetzung macht, beschränken sich deutsche Hochschulen, die das Recht auf Freiheit von Forschung und Lehre zu beachten haben, durchweg auf Empfehlungen. Die Max-Planck-Gesellschaft unterstützt das Open-Access-Publizieren durch die Übernahme anfallender Autorengebühren. Ein weiterer Anreiz könnte die Berücksichtigung von Open-Access-Publikationen als Kriterium der Mittelverteilung innerhalb einer Hochschule oder Forschungseinrichtung sein.

Die genannte Umfrage unter Open-Access-Autor(inn)en hat gezeigt, dass diese einen schnellen Publikationspro-

zess sehr schätzen. Bei Open-Access-Zeitschriften entfallen die Zeiten für den Druck und das Binden der Hefte. Eine weitere Beschleunigung wird erreicht, wenn die Hefstruktur aufgelöst und jeder einzelne Beitrag unmittelbar nach der Fertigstellung publiziert wird. Außerdem werden bei diversen Open-Access-Zeitschriften – zum Beispiel der renommierten Zeitschrift „Atmospheric Chemistry and Physics“ des Copernicus-Verlages – öffentliche Begutachtungsverfahren praktiziert, bei denen bereits die Erstfassung ins Netz gestellt wird [11], während herkömmliche Verfahren oft sehr viel Zeit kosten. Insgesamt haben Open-Access-Zeitschriften damit einen deutlichen Geschwindigkeitsvorteil, der von immerhin der Hälfte der befragten Autor(inn)en positiv wahrgenommen wird.

Anders als Zeitschriften, die parallel gedruckt und elektronisch erscheinen, können rein elektronische die speziellen Möglichkeiten dieser Publikationsform ausnutzen, zum Beispiel Links auf verwandte Literatur, sonstige Internet-Quellen und Primärdaten setzen sowie multimediale und interaktive Elemente einbinden. Überraschenderweise wurden Kommentier- und Aktualisierungsmöglichkeiten von den befragten Open-Access-Autorinnen und Autoren nicht sehr hoch als Vorteil bewertet; es könnte aber eine Frage der Zeit sein, bis sich Web-2.0-Elemente aller Art etabliert haben und aus modernen Zeitschriften nicht mehr wegzudenken sind.

Ob Open-Access-Artikel häufiger zitiert werden als konventionell publizierte, ist eine kontrovers diskutierte Frage [12]. Vor dem Hintergrund, dass Open-Access-Befürworterinnen und Befürworter gerade deren erhöhte Sichtbarkeit und Verbreitung betonen, überraschte bei der Autorenbefragung, dass nur 6,37 % subjektiv den Eindruck hatten, besonders häufig zitiert worden zu sein.

Unbestritten ist aber, dass Open-Access-Artikel schneller und leichter zugreifbar sind als kommerziell vertriebene, und zwar auch für benachteiligte Personengruppen. Das sind Wissenschaftlerinnen und Wissenschaftler in ärmeren Ländern, aber auch im eigenen Land wie z. B. Arbeitslose und andere, die keinen Zugang zu universitären Netzwerken haben, sowie im Bereich der Medizin praktizierende Ärzte. Für die Ausbildung ist der freie Zugang zu Forschungsliteratur ebenfalls förderlich, außerdem zur Verbreitung von wissenschaftlichen Ergebnissen in der Öffentlichkeit. Wenn also unter Umständen nur wenig persönlicher Nutzen beim Open-Access-Publizieren festzustellen ist, so ist der kollektive Nutzen unbestritten hoch.

Schließlich spricht der Erfolg für sich: Die Autorinnen und Autoren, die bereits mehr als 10 Open-Access-Aufsätze veröffentlicht hatten, sahen deutlich mehr Nutzen darin als der Durchschnitt und antworteten durchweg positiver auf alle Fragen. Wer sich einmal zum Open-Access-Publizieren durchgerungen hat, wird sehr wahrscheinlich wieder zu dieser Publikationsform zurückkehren.

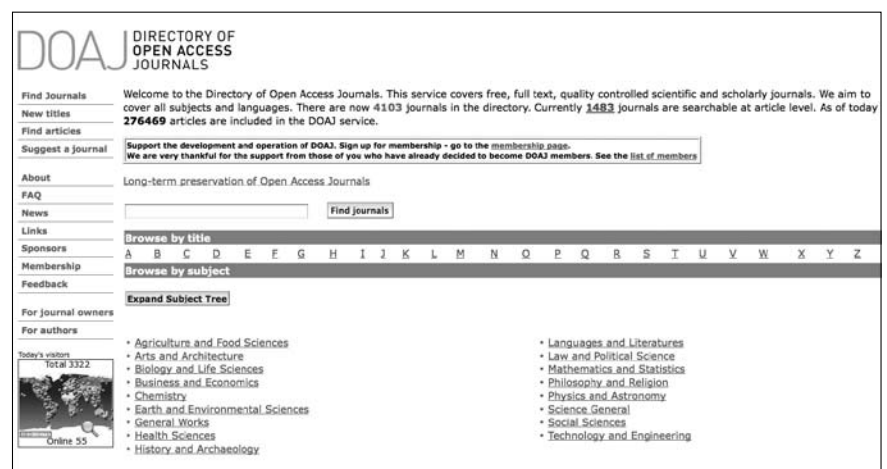


Abb. 3: Directory of Open Access Journals (URL: <http://www.doaj.org/>)



## Literatur

- [1] SCHMIDT, B.: *Auf dem „goldenen“ Weg? Alternative Geschäftsmodelle für Open-Access-Primärpublikationen.* In: Zeitschrift für Bibliothekswesen und Bibliographie 54. H. 4-5, 2007, S. 177–182. <http://eprints.rclis.org/archive/00012048/>
- [2] SIETMANN, R.: *Artikelbeschleunigung: neuer Schub für Open Access vom CERN.* In: c't, Nr. 16, 2008, S. 38–39.
- [3] *Elektronische Zeitschriftenbibliothek, Universitätsbibliothek Regensburg.* <http://www.bibliothek.uni-regensburg.de/ezeit/fl.phtml?bibid=UBR>
- [4] <http://www.ulrichsweb.com>
- [5] [http://www.thomsonreuters.com/business\\_units/scientific/](http://www.thomsonreuters.com/business_units/scientific/)
- [6] *Directory of Open Access Journals.* <http://www.doaj.org>
- [7] <http://www.openj-gate.com>
- [8] *Red de Revistas Científicas de América Latina y el Caribe, España y Portugal, Sistema de Información Científica Redalyc.* <http://www.redalyc.org>
- [9] WEISHAUP, K.: *Der freie Zugang zum Wissen: auf dem Weg, aber noch nicht am Ziel! Erste Ergebnisse einer Studie zur Akzeptanz von Open-Access-Zeitschriften.* Gelsenkirchen: Inst. Arbeit und Technik. Forschung Aktuell, Nr. 08/2008. <http://www.iat.eu/forschung-aktuell/2008/fa2008-08.pdf> sowie WEISHAUP, K.: *Vorbild Harvard in Deutschland unerreichbar? Oder: Wie lässt sich die Akzeptanz von Open-Access-Zeitschriften verbessern?* In: Ockenfeld, Marlies (Hrsg.): *Verfügbarkeit von Informationen: 30. Online-Tagung der DGI / 60. Jahrestagung der DGI, Frankfurt am Main, 15. – 17. Oktober 2008; Proceedings.* Frankfurt: DGI, S. 87–96. <http://www.iat.eu/aktuell/veroeff/2008/weishaupt06.pdf>
- [10] HERB, U.: *Alte Hüte und neue Konzepte: Qualitätssicherung, Qualitätsmessung und Zitationshäufigkeiten.* In: Telepolis, 08.11.2006. <http://www.heise.de/tp/r4/artikel/23/23829/1.html>
- [11] PÖSCHL, U.: *Mehr Transparenz und Effizienz: interaktives Open Access Publizieren und gemeinschaftliche Fachbegutachtung.* In: *Forschung & Lehre* 06, 2007, S. 334–335. [http://www.atmospheric-chemistry-and-physics.net/pr\\_acp\\_mehr\\_transparenz\\_und\\_effizienz.pdf](http://www.atmospheric-chemistry-and-physics.net/pr_acp_mehr_transparenz_und_effizienz.pdf)
- [12] Eine Untersuchung mit positivem Ergebnis wird beschrieben in: HARNAD, S. / BRODY, T.: *Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals.* In: *D-Lib Magazine* 10, no. 6, 2004. <http://www.dlib.org/dlib/june04/harnad/06harnad.html>  
Die gegenteilige Meinung wird dargestellt in: DAVIS, P. M. / LEWENSTEIN, B. V. / SIMON, D. H. / BOOTH, J. G. / CONNOLLY, M. J. L.: *Open access publishing, article downloads, and citations: randomised controlled trial.* In: *BMJ – British medical journal*, Nr. 568, 2008, S. 337 ff. [http://www.bmj.com/cgi/content/full/337/jul31\\_1/a568?q=rss\\_home](http://www.bmj.com/cgi/content/full/337/jul31_1/a568?q=rss_home)

# Auf dem goldenen Weg zum edoc-Server – Open-Access-Zeitschriften, Schriftenreihen und Sammelbände an der Humboldt-Universität

Uwe Müller | u.mueller@cms.hu-berlin.de

*Der edoc-Server der Humboldt-Universität bildet nicht nur das Institutional Repository der Hochschule im Sinne des grünen Weges zu Open Access. Er stellt auch eine Veröffentlichungsplattform für Erstveröffentlichungen unterschiedlicher Art dar. Dazu zählen vor allem periodisch erscheinende Publikationen wie elektronische Zeitschriften, Schriftenreihen und Jahrbücher, aber auch einmalige Publikationsvorhaben wie Konferenzbände oder Sammelbände. Der Beitrag stellt das derzeitige Angebot dar, das mit dem edoc-Server für Herausgeber bereitgehalten wird und geht auf die technologische und organisatorische Unterstützung ein, die sie bei der Umsetzung in Anspruch nehmen können.*

*Open-Access-Zeitschrift, Sammelband, Schriftenreihe, XML, Workflow, Hosting*

Der freie und uneingeschränkte Zugang zu wissenschaftlichen Publikationen lässt sich über verschiedene Wege erreichen. Eine nahe liegende Möglichkeit besteht darin, bestehende Publikationsformen in Open-Access-basierte Geschäftsmodelle zu überführen, die entstehenden Veröffentlichungen also online und kostenfrei anzubieten, während der organisatorische Publikationsprozess weitgehend unverändert bleibt. Der entscheidende Vorteil dieses so genannten goldenen Weges zu Open Access (vgl. [1]) besteht darin, dass etablierte und erprobte Organisationsformen übernommen werden können, wodurch die grundsätzlichen Anforderungen an das wissenschaftliche Publizieren ebenso erfüllt werden wie bei traditionellen und kostenpflichtigen Publikationen. Dies betrifft beispielsweise die Verfahren der inhaltlichen Qualitätssicherung (siehe [2]), aber auch die üblichen Konventionen für die Zitierung von Veröffentlichungen.

Bekannteste Ausprägung für den goldenen Weg sind die Open-Access-Zeitschriften, für die es inzwischen in allen Wissenschaftsdisziplinen einschlägige Beispiele gibt und die vergleichbaren Subskriptionszeitschriften in Bezug auf Renommee und wissenschaftliche Qualität im Allgemeinen in keiner Weise nachstehen. Aber auch andere periodisch erscheinende Veröffentlichungen wie Schriftenreihen und Jahrbücher oder einmalige Publikationsvorhaben wie Konferenzbände und Sammelbände erscheinen inzwischen vermehrt elektronisch und in Form von Open Access.

## Veröffentlichen ...

Als allgemeiner Hochschulpublikationsserver stellt der edoc-Server der Humboldt-Universität neben den klassischen Hochschulschriften wie Dissertationen, Habilitationsschriften und Studienabschlussarbeiten sowie Sekundärveröffentlichungen im Rahmen des grünen Weges zu Open Access – so genannte Postprints – auch eine Vielzahl mehrteiliger Publikationsvorhaben bereit. Sie erscheinen unter der (Mit-)Herausgesellschaft von Universitätsmitgliedern und sind klar dem goldenen Weg zu Open Access zuzuordnen. Diese Veröffentlichungen bilden sowohl in Bezug auf die inhaltliche Ausrichtung als auch hinsichtlich ihrer Entstehungsgeschichte, Erscheinungsweise und organisatorischer Rahmenbedingungen ein breites Spektrum ab. Derzeit werden auf dem edoc-Server Open-Access-Zeitschriften aus den Fachgebieten Geschichte, Kunstgeschichte, Skandinavistik, Japanologie, Bibliotheks- und Informationswissenschaft, Informatik, Mathematik und Museumskunde veröffentlicht. Hinzu kommen zahlreiche Sammelbände, Schriftenreihen und Konferenzbände, unter anderem aus den Wirtschaftswissenschaften, den Agrarwissenschaften, der Theologie, der Geografie und der Ethnologie. Unter den veröffentlichten Journalen sind solche, die erst vor relativ kurzer Zeit unmittelbar als Open-Access-Zeitschriften gegründet wurden, und andere, die zunächst ausschließlich gedruckt und teilweise in einem Verlag erschienen, bevor sie im Verlaufe ihres Bestehens in die elektronische Form überführt wurden. Während die meisten



Zeitschriften auf dem edoc-Server ausschließlich elektronisch erscheinen, wird beispielsweise für das *cms-journal* parallel zur Online-Fassung eine gedruckte Ausgabe verlegt.

den Titel, die Namen der Autoren und mögliche Seitenangaben aber ggf. auch Schlagwörter und Abstracts. Je nach Besonderheiten der einzelnen Publikationen lassen sich die Eingabemasken

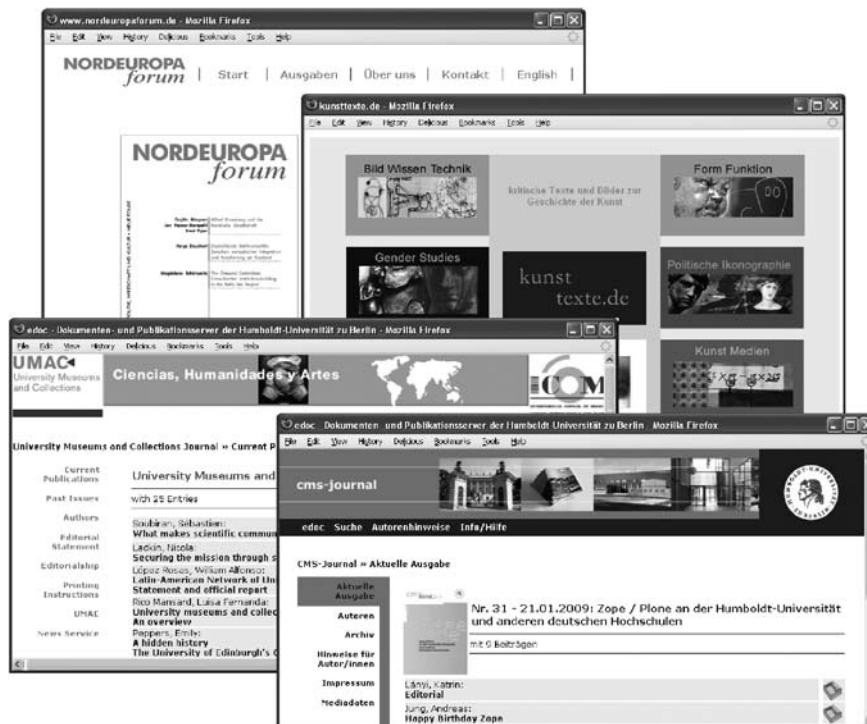


Abb. 1: eine Auswahl der Open-Access-Zeitschriften, die derzeit auf dem edoc-Server veröffentlicht werden

Die Veröffentlichung elektronischer Zeitschriften und anderer periodischer bzw. mehrteiliger Publikationsvorhaben auf dem edoc-Server beinhaltet in erster Linie die Bereitstellung einer Publikationsplattform. Auf ihr werden die einzelnen Ausgaben und Beiträge dauerhaft online zur Verfügung gestellt und als Volltext abrufbar gehalten. Der Service gegenüber dem Herausgeber einer solchen Publikation schließt jedoch darüber hinaus eine umfassende technologische und organisatorische Unterstützung des gesamten Publikationsprozesses mit ein.

Im einfachsten Fall können für ein Publikationsvorhaben die Mitglieder der Redaktion über eine Webschnittstelle die einzelnen Beiträge eigenständig einstellen, verwalten, anordnen und freischalten – und damit formal veröffentlichen. Das umfasst neben dem Upload der fertigen Dateien – zumeist in Form archivierbarer PDF/A-Dokumente – auch die Erfassung von Metadaten zu den einzelnen Beiträgen, also etwa

konfigurieren und beispielsweise durch weitere Datenfelder ergänzen.

Auch die Gestalt, mit der sich beispielsweise eine elektronische Zeitschrift ihren Lesern präsentieren soll, lässt sich den Wünschen der Herausgeber anpassen. Das betrifft nicht nur das Layout des Webauftritts und die Anordnung der einzelnen Elemente, sondern beispielsweise auch die Recherchestrategie, die sich aus den jeweils angebotenen Navigationsmöglichkeiten ergibt. Zusätzlich zu den klassischen Suchfunktionen können dabei einzelne Beiträge anhand unterschiedlicher Kategorien gezielt und intuitiv angesteuert und aufgefunden werden – etwa über eine Auflistung aller verfügbaren Ausgaben einer Zeitschrift, einen Autoren-Index oder eine Schlagwortliste (siehe Abb. 2). Je nach Bedarf erhalten Periodika, insbesondere Zeitschriften, natürlich auch ihre eigenen Domains, wie beispielsweise im Falle von [www.kunsttexte.de](http://www.kunsttexte.de) und [www.nordeuropaforum.de](http://www.nordeuropaforum.de).



Abb. 2: Im Nordeuropaforum können nicht nur die einzelnen Zeitschriftenausgaben angesteuert werden; es ist auch eine Recherche nach Autoren und Schlagwörtern möglich.

Auf dem edoc-Server publizierte Zeitschriften, Schriftenreihen und Sammelbände sind nicht nur unmittelbar im Webangebot sichtbar und über die bekannten Suchmaschinen auffindbar. Sie sind auch in übergreifende Recherchedienste eingebunden, sodass die Sichtbarkeit derart verfügbar gemachter Publikationen beispielsweise gegenüber einfachen Webseiten deutlich erhöht ist (vgl. auch [3]).

## ... und Publizieren mit XML

Abgesehen von dem Basisdienst, mit dem Zeitschriften und andere Publikationsvorhaben bequem per Webschnittstelle verwaltet werden können und in eine professionelle Veröffentlichungsplattform eingebunden sind, können Herausgeber auch bei der technischen Erstellung der eigentlichen Publikationen, also vor allem bei der Erzeugung des Satzspiegels, unterstützt werden. Dazu stehen mehrere Ansätze zur Verfügung, mit denen mithilfe unterschiedlicher Desktop-Publishing-Systeme die einzelnen Beiträge einer Publikation in ein einheitliches Layout überführt werden können. Unabhängig von der Wahl des konkreten Systems – im Einsatz befinden sich derzeit QuarkXPress, Adobe Framemaker und Adobe Indesign – geht es dabei nicht ausschließlich um die Generierung eines ansprechenden Druckdesigns. Durch die Nutzung von XML als zugrunde liegende Beschreibungssprache entsteht für jeden Beitrag

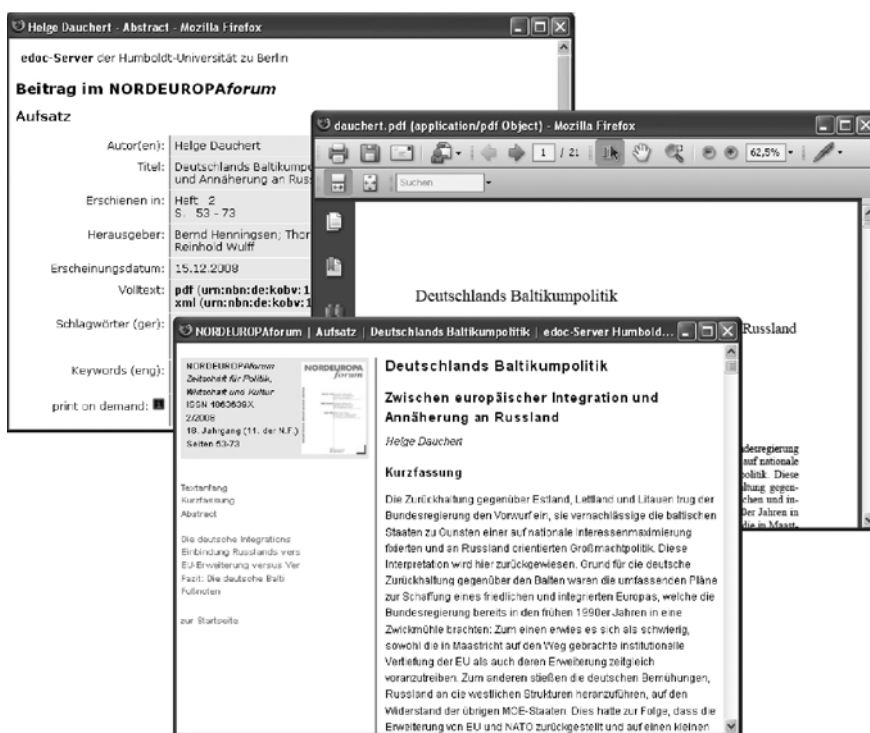


Abb. 3: Startseite, PDF-Fassung und dynamisch erzeugte XML-Fassung eines Artikels in der Zeitschrift „Nordeuropaforum“

mit der Erstellung der für den Druck optimierten PDF-Datei parallel und ohne zusätzlichen manuellen Aufwand eine HTML-Version, die sich für die Lektüre am Bildschirm wesentlich besser eignet als die PDF-Fassung. Die Nutzungsqualität der Publikation wird also entscheidend erhöht. Dieser integrierte XML-basierte Dokumenterstellungsprozess setzt voraus, dass charakteristische Strukturmerkmale – etwa Überschriften, Legenden und Fußnoten – aber auch inhaltsbezogene Hervorhebungen – z. B. Fachwörter, Formeln oder Ortsnamen – explizit gekennzeichnet werden. Dies kann durch die Redaktion geschehen, lässt sich aber auch bereits durch die Autoren bewerkstelligen, denen zu diesem Zweck entsprechende Dokumentvorlagen zur Verfügung gestellt werden. Auch wenn die Verwendung derartiger Hilfsmittel während der Texterstellung auch unabhängig von der späteren Weiterverarbeitung den Autoren klar zum Vorteil gereicht, befürchten Herausgeber zuweilen, sie könnten ihre potentiellen Textlieferanten überfordern oder abschrecken, wenn sie die Verwendung einer Dokumentvorlage vorschreiben. Zumindest als optionales Angebot sollte eine Dokumentvorlage, die für die

betreffende Zeitschrift oder Schriftenreihe angepasst ist, jedoch vorgehalten werden, wenn ein XML-basierter Dokumenterstellungsprozess vorgesehen ist.

Selbstverständlich ist der Aufwand zur Einrichtung eines solchen integrierten Publikationsprozesses einschließlich der Umsetzung individueller Layoutanforderungen sowie der Schulung der Mitarbeiter durchaus beachtlich. Allerdings werden die Vorteile des elektronischen Publizierens auch erst durch diesen XML-basierten Ansatz vollständig wirksam. Besonders lohnenswert ist der Einsatz eines integrierten Dokumenterstellungsprozesses für häufig erscheinende und umfangreiche Zeitschriften, da der einmalig anfallende Anfangsaufwand durch die Zeitersparnis im Regelbetrieb schnell aufgewogen wird.

## Workflow-Unterstützung

Die Herausgabe einer Zeitschrift oder einer anderen periodisch erscheinenden bzw. mehrteiligen Publikation ist in einen umfassenden Publikationsprozess eingebettet, der weit über das Einstellen und Freischalten fertiger Dokumente hinaus-

geht. Mit der Herstellung eines anspruchsvollen und für unterschiedliche Ausgabearten geeigneten Layouts wurde im vorherigen Abschnitt bereits ein wesentlicher Bestandteil dieses Prozesses genannt. Hinzu kommt der gesamte organisatorische Teil, der unter anderem die Kommunikation mit den Autoren sowie die inhaltliche Qualitätssicherung durch jeweils zugeordnete Gutachter einschließt. Um diesen komplexen und für jeden eingereichten Beitrag durchzuführenden Prozess zu strukturieren und handhabbar zu halten, existieren verschiedene webbasierte Manuskriptverwaltungssysteme, deren Kernfunktion die Organisation des Begutachtungsprozesses ist. Herausgeber, die auf dem edoc-Server publizieren, können eine Installation von *Open Journal Systems*, der am weitesten verbreiteten Software auf diesem Gebiet, nutzen. Damit können nicht nur Autoren ihre Manuskripte unmittelbar online einreichen und später ggf. verändern. Auch die Redaktion, die Gutachter und alle technisch involvierten Personen haben direkten Zugriff auf die für sie relevanten Vorgänge und Dokumentversionen der einzelnen Beiträge, die dort auch dauerhaft archiviert werden. Die Arbeit mit dem System erfolgt ausschließlich via Webbrowser. Über nächste Schritte, bevorstehende Fristen und Ähnliches werden die jeweils Beteiligten durch automatisch versandte E-Mails informiert. Beispielsweise wird die Redaktion über neu eingegangene Manuskripte ebenso unterrichtet wie Gutachter über die Bitte, ein Manuskript zu begutachten. Die verantwortlichen Redaktionsmitglieder können dabei jederzeit einen aktuellen Überblick über den Status aller Vorgänge erlangen und selbstverständlich auch in die laufenden Prozesse eingreifen. Die manuelle Organisation des Publikationsprozesses kann somit wesentlich effizienter und zeitsparender gestaltet werden. Die Daten der schließlich akzeptierten und zur Publikation bestimmten Beiträge können anschließend ohne Probleme auf den edoc-Server übernommen werden.

## Und sonst?

Die Publikationsplattform des edoc-Servers besteht nur auf den ersten Blick ausschließlich aus Online-Speicherplatz für elektronische Veröffentlichungen. Neben den benannten Möglichkeiten zur Unterstützung einer XML-basierten Dokumenterstellung und des organisatorischen Publikationsprozesses bietet der edoc-Server weitere Mehrwerte aus Sicht von Herausgebern elektronischer Publikationsvorhaben. Zum einen ist durch die Bereitstellung standardisierter Schnittstellen und die nationalen und internationalen Kooperationen des edoc-Servers eine gute Sichtbarkeit der einzelnen Veröffentlichungen gegeben, weil sie durch automatisierten Austausch der Metadaten in zahlreichen Rechercheportalen und Fachdatenbanken nachgewiesen werden. Der edoc-Server garantiert nicht nur die langfristige Verfügbarkeit der Publikationen, sondern stellt durch die Vergabe so genannter *Persistent Identifiers* in Form einer URN für jeden einzelnen Beitrag sicher, dass er auch bei ggf. veränderter

Domain oder einer technischen Umstrukturierung des Servers auf identische Weise zitierbar bleibt.

Für alle Publikationsvorhaben werden detaillierte und auf einzelne Beiträge bezogene Zugriffsstatistiken geführt, die auf Wunsch auch online zur Verfügung gestellt werden und beispielsweise Aufschluss darüber geben, welche Beiträge besonders häufig aufgerufen worden sind. Der edoc-Server verfügt über eine Print-on-Demand-Funktion, die auch auf der Ebene einzelner Beiträge mehrteiliger Publikationen komfortabel genutzt werden kann – etwa, indem ausgewählte Artikel aus einer Zeitschrift als individuell zusammengestellter Reader bestellt werden können.

Die Metadaten einer Veröffentlichung dienen nicht nur der Darstellung und der besseren Auffindbarkeit, sondern dem interessierten Leser ggf. auch als Grundlage zur eigenen Weiterarbeit – insbesondere beim Verfassen eigener Publikationen. Um die bibliografischen Angaben zu einer Publikation nahtlos in eine eigene Literaturverwaltung zu übernehmen, kann die

Exportfunktion des edoc-Servers verwendet werden, die die Formate BibTex und EndNote bereitstellt. Darüber hinaus besteht eine Anbindung an gebräuchliche Social-Bookmarking- bzw. Tagging-Systeme.

Schließlich lassen sich die einzelnen Publikationsvorhaben auch so ausstatten, dass Autoren selbst wählen können, welche Rechte Dritte an ihren Veröffentlichungen haben sollen. Dies geschieht mithilfe der so genannten Creative-Commons-Lizenzen, die in dem vorliegenden Heft an anderer Stelle detailliert beschrieben werden [4].

## Fazit

Der edoc-Server ist nicht nur Hochschulschriften- und Dokumentenserver der Universität, sondern beherbergt auch eine Vielzahl regelmäßig erscheinender Publikationsvorhaben, die von den jeweiligen Herausgebern und Redaktionsgruppen weitgehend eigenständig und unabhängig verwaltet werden. Die freie Verfügbarkeit aller hier publizierten Inhalte im Sinne von Open Access ist die grundlegende Maxime, die mit dem Betrieb des edoc-Servers verfolgt wird. Wer die Herausgabe einer Zeitschrift, einer Schriftenreihe, eines Sammelbandes, eines Jahrbuchs oder eines Konferenzbandes plant, sollte sich über die Möglichkeiten und Konditionen informieren, die der edoc-Server dafür bietet. Aber auch, wenn bereits bestehende Publikationsvorhaben professionalisiert, erstmals elektronisch angeboten oder in ein Open-Access-basiertes Geschäftsmodell überführt werden sollen, ist der edoc-Server sicherlich eine gute Wahl.

## Literatur

- [1] WEISHAUPT, K.: Überblick zum goldenen Weg ... *cms-journal* 32, 2009, S.24–28.
- [2] MÜLLER, U.: Peer Review bei Open-Access-Zeitschriften. *cms-journal* 32, 2009, S. 33–37.
- [3] MALITZ, R.: Open Access – Verfügbar ist noch nicht präsent. *cms-journal* 32, 2009, S. 70–73.
- [4] KIATT-KAFEMANN, M.: Creative Commons. *cms-journal* 32, 2009, S. 81–86.

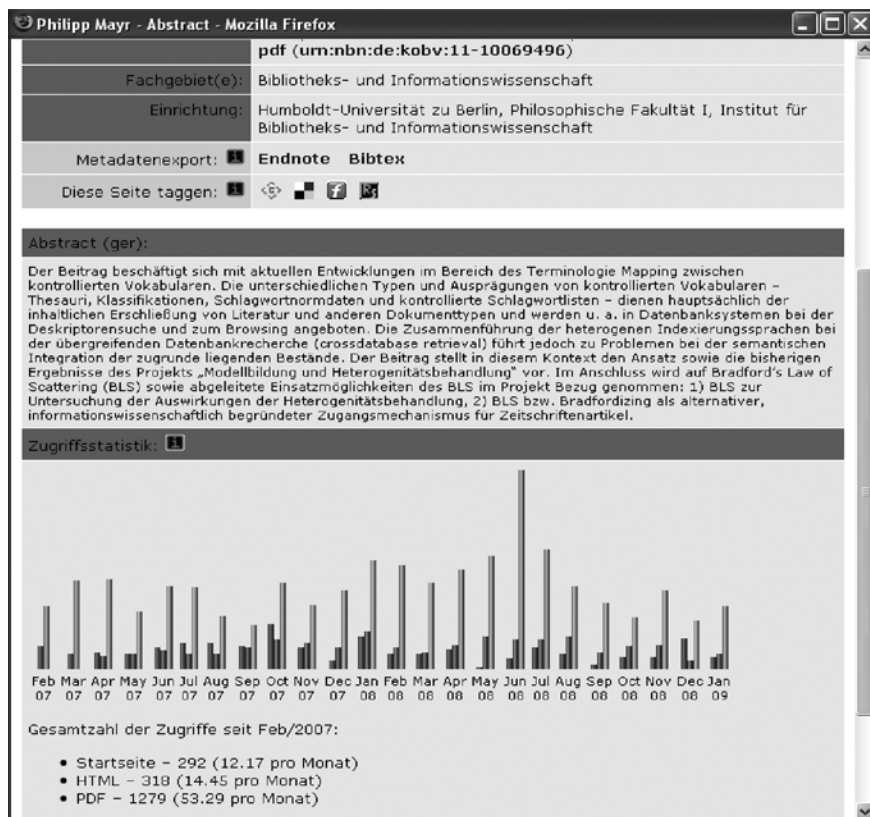


Abb. 4: Metadatenexport, Tagging-Funktionen und detaillierte Zugriffsstatistik für einen Beitrag eines Sammelbandes

# Peer Review bei Open-Access-Zeitschriften

Uwe Müller | u.mueller@cms.hu-berlin.de

## Qualitätssicherung beim Publizieren

Die Qualitätssicherung gilt als bedeutendes Merkmal wissenschaftlichen Publizierens. Dazu zählt die sorgfältige Prüfung der Manuskripte vor deren Einreichung durch die Verfasser selbstverständlich ebenso wie das kritische Hinterfragen veröffentlichter Arbeiten und deren Inhalte durch die Rezipienten. Der Kern der Qualitätssicherung wird jedoch durch einen Begutachtungsprozess gebildet, der durch den jeweiligen Herausgeber einer Publikation organisiert wird und sicherstellen soll, dass nur solche Arbeiten überhaupt veröffentlicht werden, die bestimmte Qualitätskriterien erfüllen. Dazu werden Fachgutachter hinzugezogen und damit beauftragt, die wissenschaftliche Relevanz des Manuskripts und den Wahrheitsgehalt bzw. die Verlässlichkeit des Inhalts ebenso zu bewerten wie Nachvollziehbarkeit, Konsistenz und Korrektheit der verwendeten Ansätze und Methoden sowie die zutreffende Einordnung der Arbeit in den fachlichen Kontext. Gegenstand der Überprüfung – und ggf. von Verbesserungsvorschlägen seitens der Gutachter – sind darüber hinaus in der Regel auch Darstellung, Verständlichkeit, Prägnanz, Form und Stil des betreffenden Manuskripts. Besonderer Stellenwert genießt dieses so genannte *Peer Review* – als *Peers* werden die jeweiligen Gutachter bezeichnet, die zumeist ebenso wie die Autoren aktive Wissenschaftler sind – bei wissenschaftlichen Zeitschriften. *Peer Review* dient dabei nicht nur als Maßnahme der Filterung, die die Literatursuche und -bewertung seitens potentieller Rezipienten

erheblich vereinfacht. Zu den wichtigsten Funktionen der Qualitätskontrolle zählen auch die Steigerung der Reputation der Autoren sowie die Sicherung des Renommées der Zeitschrift. Denn selbstverständlich deuten Publikationen, die eine kritische Qualitätsbewertung erfolgreich überstanden haben, eher auf die Leistungsfähigkeit eines Wissenschaftlers hin, als Veröffentlichungen, die ein solches „Qualitätssiegel“ nicht tragen. Und insofern ist es – abgesehen vom tatsächlichen Inhalt eines wissenschaftlichen Artikels – unter anderem für dessen Glaubwürdigkeit und Akzeptanz in der Fachwelt durchaus nicht gleichgültig, in welcher Zeitschrift er erschienen ist.

Dies gilt auch für Open-Access-Zeitschriften, denen von vielen Wissenschaftlern – Autoren wie Rezipienten – oftmals per se eine geringere inhaltliche Qualität unterstellt wird, und die damit häufig mit einem geringen Renommee in Verbindung gebracht werden. Das liegt zum einen sicherlich an der zumeist noch relativ kurzen Historie dieser Periodika – das Renommee einer Zeitschrift ist in der Regel nicht unmittelbar seit ihrer Gründung vorhanden und muss sich sozusagen erst „erweisen“. Auf der anderen Seite gilt aber die Vermutung, dass Open-Access-Zeitschriften an das interne *Peer-Review*-Verfahren geringere Maßstäbe anlegen könnten als vergleichbare Subskriptionszeitschriften. Begründet wird dieser Verdacht unter anderem damit, dass das Open-Access-basierte Geschäftsmodell mit seinen Autorengebühren einerseits und der fehlenden „Macht“ der zahlenden Abonnenten andererseits den Anreiz dazu bietet, auch minderwertige Beiträge zu publizieren. Unab-

*Die Qualitätssicherung gilt als konstitutionell für das wissenschaftliche Publizieren und die Wissenschaft insgesamt. Sie hat sich in Gestalt organisierter gegenseitiger Kritik im Rahmen des so genannten Peer Review seit Jahrhunderten etabliert und durchgesetzt. Unter anderem wegen ihrer abweichenden Geschäftsmodelle stehen Open-Access-Zeitschriften grundsätzlich im Verdacht, geringere Maßstäbe an die interne Qualitätssicherung zu setzen. Eine umfassende Studie, deren wichtigste Ergebnisse in diesem Beitrag beschrieben werden, widerlegt diese Vermutung.*

*Open-Access-Zeitschrift, Peer Review, Begutachtung, Qualität,*



hängig davon, in welchem Maße diese Annahmen zutreffend sind, speisen sich Vorbehalte gegenüber Open Access und eine vergleichsweise geringe Akzeptanz von Open-Access-Zeitschriften unter Wissenschaftlern daraus, wie entsprechende Untersuchungen nahelegen (siehe z. B. [1]). Es leuchtet ein, dass derlei Zuschreibungen der Verbreitung von Open Access insgesamt nicht zum Vorteil reichen, wie auch der Open-Access-Vordenker Stevan Harnad konstatiert, wenn er schreibt: "The refereed journal literature needs to be freed from both paper and its costs, but not from peer review, whose 'invisible hand' is what maintains its quality." [2]

## Peer Review unter der Lupe

Abgesehen davon, dass inzwischen viele Open-Access-Zeitschriften durch ihren auf Zitationen basierenden und damit die tatsächliche Wertschätzung durch andere Wissenschaftler repräsentierenden *Journal Impact Factor* als ihren traditionellen Pendanten zumindest gleichwertig und in einigen Fällen auch als überlegen einzustufen sind – als Beispiele seien hier die Zeitschrift *Atmospheric Chemistry and Physics* (ACP) und die Journale der *Public Library of Science* (PLOS) genannt – sollte in einer umfangreichen Studie ermittelt werden, wie Peer-Review-Verfahren unter heutigen Open-Access-Zeitschriften tatsächlich aussehen<sup>1</sup>. Dazu wurden im Frühjahr 2008 die Herausgeber von mehr als 3.000 Zeitschriften befragt, die im weltweit umfassendsten Verzeichnis wissenschaftlicher OpenAccess-Zeitschriften, dem im schwedischen Lund geführten *Directory of Open Access Journals* (DOAJ), registriert sind. In Bezug auf die tatsächlich existierenden Open-Access-Zeitschriften ist die Stichprobe damit sehr umfassend gewählt und zudem international und interdisziplinär angelegt.

Zentraler Gegenstand der mit einem Online-Fragebogen realisierten Umfrage waren Art und Beschaffenheit der jeweils verwendeten Peer-Review-Verfahren. Im

Einzelnen ging es dabei unter anderem um das Einreichungsverfahren, das Verhältnis zwischen Herausgebern, Redaktion und Gutachtern, die Auswahl konkreter Gutachter, die gegenseitige Anonymität von Gutachtern und Autoren, die Transparenz des Verfahrens und den Informationsfluss – also die Frage, wer wann Zugang zu welchen Informationen erhält. Daneben wurde auch untersucht, inwieweit mögliche Interessenkonflikte aufseiten von Autoren und Gutachtern betrachtet werden, ob eine Qualitätskontrolle von Gutachten stattfindet und wie hoch der technische Unterstützungsgrad für die am Publikationsprozess beteiligten Akteure ist – also wie die Kommunikation abläuft und ob ein Manuskriptverwaltungssystem zum Einsatz kommt.

Insgesamt wurde der recht umfangreiche Fragebogen durch Verantwortliche von mehr als 1.200 Open-Access-Zeitschriften ausgefüllt. Das entspricht einer Rücklaufquote von etwa 40 %. Die Ergebnisse der Untersuchung legen nahe, dass Begutachtungsverfahren zur Qualitätssicherung in der überwiegenden Mehrzahl der Open-Access-Zeitschriften verwendet werden. In der Auswertung wurde vor allem deutlich, dass die Gestalt von Peer-Review-Verfahren wesentlich vom Fachgebiet der Zeitschrift und der Art des Verlegers abhängt. Darüber hinaus konnten Korrelationen zwischen einzelnen Verfahrenseigenschaften untereinander und in Bezug auf externe Qualitätsindikatoren für Zeitschriften – insbesondere der Indexierung im *Science Citation Index* (SCI) – nachgewiesen werden.

## Unterschiede je nach Fachgebiet ...

Die fachbezogene Beurteilung<sup>2</sup> der erfassten Daten ergab unter anderem, dass für geistes- und sozialwissenschaftliche Zeitschriften durchschnittlich deutlich weniger Redaktionsmitglieder und Gutachter tätig sind als bei Zeitschriften im STM-Bereich<sup>3</sup> und dass jeder Gutachter jährlich mehr Manuskripte bewerten

muss, obwohl die Zeitschriften insgesamt weniger Beiträge verarbeiten und publizieren. Die höchsten Ablehnungsquoten haben Zeitschriften in den Wirtschafts- und den Ingenieurwissenschaften sowie in der Mathematik zu verzeichnen. Besonders deutlich werden die fachbezogenen Unterschiede beispielsweise auch bei der Frage, wie unabhängig die Begutachtungsverfahren ihrer Organisationsform nach aufgebaut sind, wer also jeweils tatsächlich als Gutachter tätig ist. Abb. 1 zeigt die relative Häufigkeit unterschiedlicher Qualitätsstufen von Peer-Review-Verfahren hinsichtlich der Fachgebiete, aus denen die einzelnen Zeitschriften kommen. Die höchsten Quoten unabhängiger Peer-Review-Verfahren – das sind solche, bei denen die Gutachter selbst nicht zum Redaktionskreis gehören, sondern unabhängig von der Zeitschrift sind (jeweils der linke Balken) – ergeben sich bei den Grundlagenfächern der Naturwissenschaften, also Chemie und Physik/Astronomie. Am geringsten ist dieser Anteil bei den Geisteswissenschaften sowie im Bereich Kunst und Architektur.

Insbesondere in den Geisteswissenschaften sind Begutachtungsprozesse in der Regel wesentlich informeller aufgebaut, als dies im STM-Bereich der Fall ist. Das zeigt sich unter anderem daran, dass sowohl die autorensseitige Möglichkeit, Gutachter vorab auszuschließen, als auch die Gelegenheit für die Verfasser, auf ablehnende Gutachten klärend zu reagieren, vor allem bei STM-Zeitschriften besteht.

Eine ähnlich klare Trennlinie hinsichtlich der großen Wissenschaftsdisziplinen lässt sich bezüglich der gegenseitigen Anonymität von Autoren und Gutachtern feststellen. Das wegen der diesbezüglichen Informationssymmetrie insgesamt als ausgewogener und fairer eingestufte *Double Blind Peer Review* (DBPR) – hierbei kennen sich Autoren und Gutachter eines Manuskripts gegenseitig nicht – findet vor allem in den Wirtschafts- und Sozialwissenschaften, aber auch in den Geisteswissenschaften Anwendung. In den Naturwissenschaften ist diese Version dagegen fast gar nicht verbreitet. Stattdessen herrscht hier das traditionelle *Single Blind Peer Review* (SBPR) vor, bei dem die Gutachter zwar die Identitäten

1 Die Untersuchung entstand im Rahmen der Dissertation des Autors, in der auch die Ergebnisse ausführlich dargestellt sind (siehe [3]).

2 Die Fachklassifikation wurde unmittelbar aus dem DOAJ übernommen und liegt daher in Englisch vor.

3 STM steht für Science, Technology and Medicine.

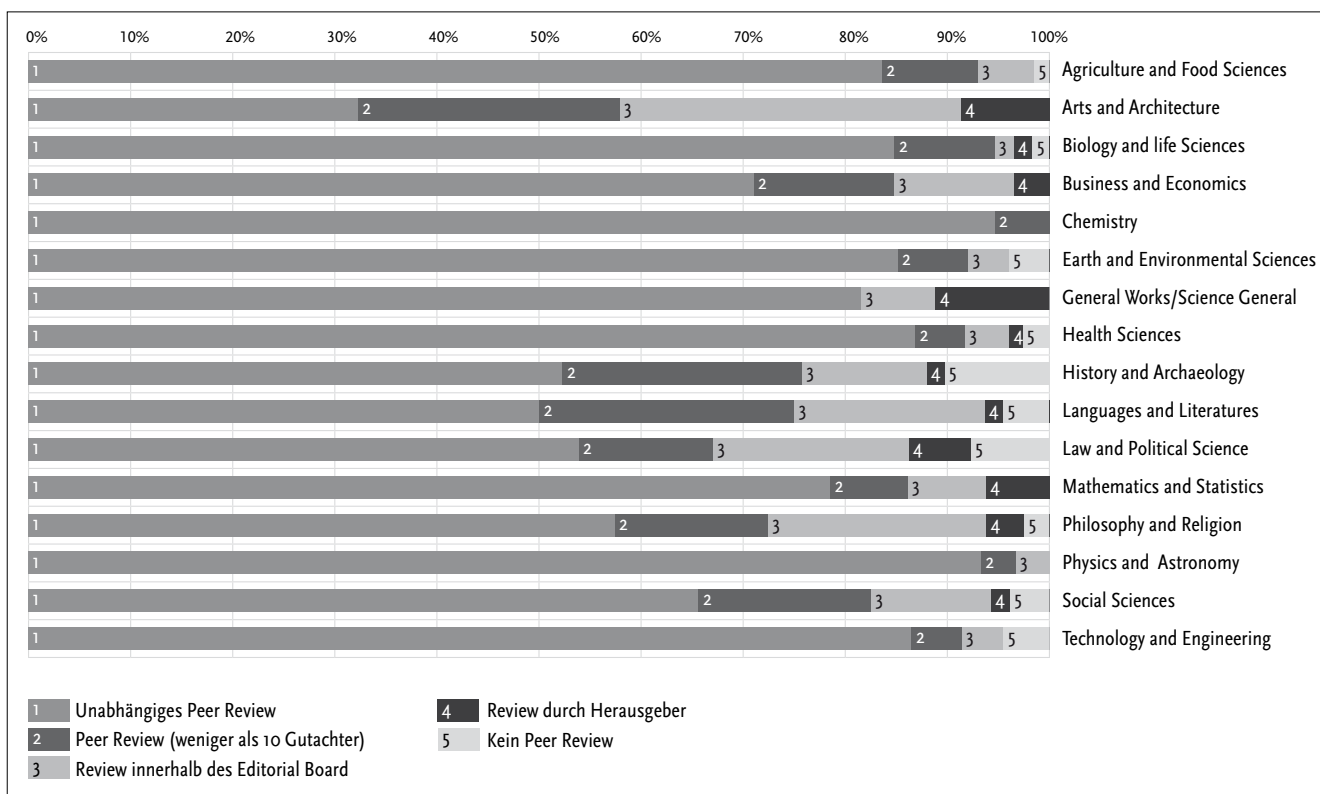


Abb. 1: Ausprägungen von Peer-Review-Verfahren nach Fachgebiet<sup>2</sup>

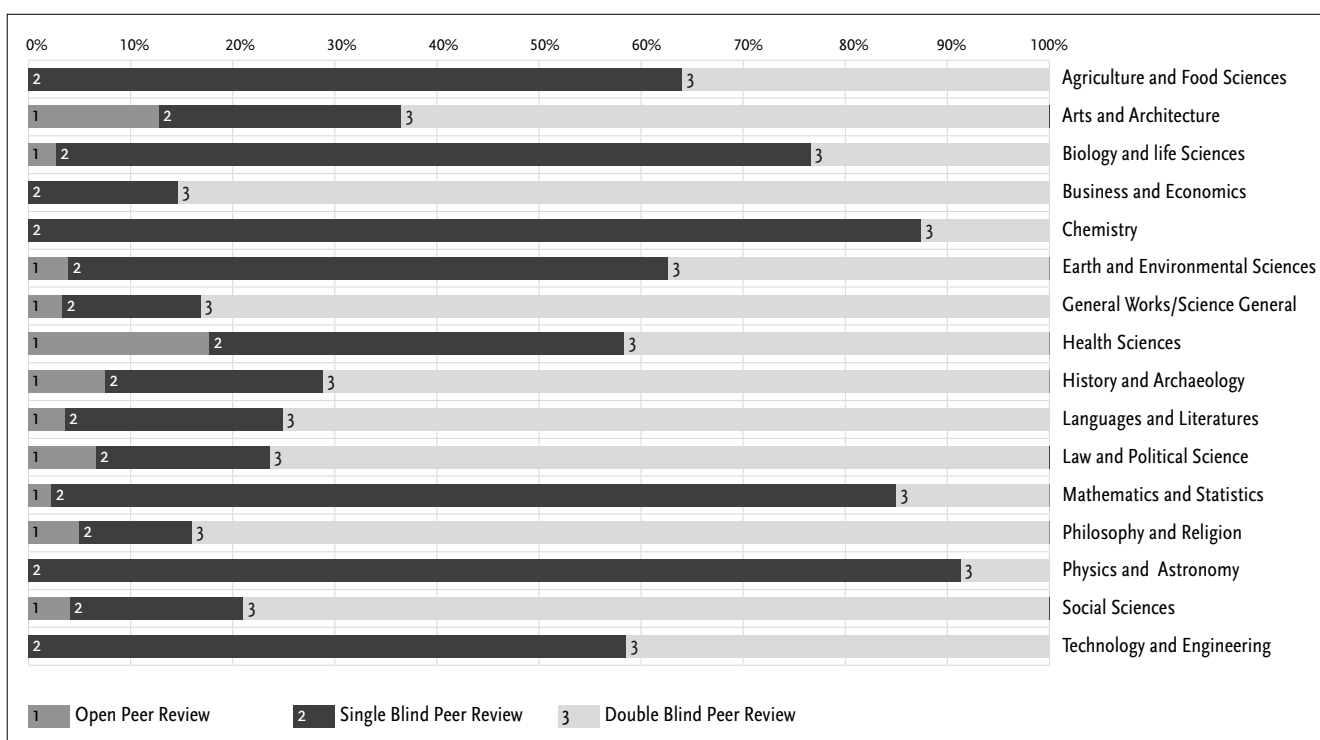


Abb. 2: Anonymität beim Peer Review nach Fachgebiet<sup>2</sup>

der Autoren kennen, die Autoren andererseits jedoch nicht wissen, wer ihr Manuskript bewertet. Die Ingenieurwissenschaften und die Medizin nehmen in dieser Hinsicht eine mittlere Position ein (siehe Abb. 2).

In gewisser Weise wiegt die häufiger eingesetzte Anonymisierung zu begutachtender Manuskripte bei den Geistes- und Sozialwissenschaften die geringe Unabhängigkeit hinsichtlich der Gutachter und den insgesamt informeller gestalteten Prozess auf. Während sich Begutachtungsprozesse bei STM-Zeitschriften insgesamt durch eine höhere Professionalisierung auszeichnen – getragen unter anderem durch mehr personelle Ressourcen, aber auch durch die stärkere Nutzung technischer Systeme zur Unterstützung des Publikationsprozesses – wird bei geistes- und sozialwissenschaftlichen Zeitschriften die Qualität des Begutachtungsverfahrens durch andere Maßnahmen – vor allem durch die Anonymisierung von Manuskripten für die Gutachter – verbessert.

## ... und Art des Verlegers

Starke Abhängigkeiten einzelner Verfahrenseigenschaften ergeben sich auch hinsichtlich der Art des Verlegers, bei dem eine Zeitschrift erscheint. Viele der untersuchten Open-Access-Zeitschriften werden nicht bei Verlagen, sondern durch hochschuleigene Fachbereiche, wissenschaftliche Fachgesellschaften, außeruniversitäre Forschungseinrichtungen oder durch informell zusammengesetzte Herausgebergruppen, beispielsweise Wissenschaftler aus unterschiedlichen Instituten, herausgegeben. Wissenschaftliche Verlage verfügen mit Abstand über die größten personellen Ressourcen zur Herausgabe einer Zeitschrift. Sie bieten im Gegensatz zu anderen Verlegerarten ihren Autoren fast immer die Möglichkeit, Gutachter auszuschließen bzw. selbst vorzuschlagen und auf negative Gutachten klärend zu antworten. Darüber hinaus haben sie hinsichtlich der Erstellung von Gutachten die genauesten Vorgaben und evaluieren sie am häufigsten. Verlage nutzen außerdem überdurchschnittlich oft aktuelle technologische Hilfsmittel –

insbesondere Manuskriptverwaltungssysteme einschließlich einer webbasierten Abgabeschnittstelle für einzureichende Manuskripte – und bieten Kommentarfunktionen für veröffentlichte Beiträge an. Dagegen geben sie im Gegensatz zu Fachgesellschaften und anderen Verlegertypen nur selten Empfehlungen hinsichtlich alternativer Publikationsmedien für Autoren, deren Manuskripte abgelehnt wurden.

Naturwissenschaftliche und medizinische Open-Access-Zeitschriften, die im SCI indexiert werden und damit einen Journal Impact Factor besitzen, werden hauptsächlich durch Fachgesellschaften, Verlage und außeruniversitäre Forschungseinrichtungen verlegt. Der Anteil derartig formal herausgehobener Zeitschriften unter den anderen Verlegertypen – Universitäten und einzelne Wissenschaftler – ist wesentlich geringer. Auch sonst zeigt sich an vielen Stellen eine deutliche Korrelation zwischen Ausprägungen des Begutachtungsverfahrens und der Indexierung im SCI. Unter anderem ist der Anteil der Zeitschriften mit einem unabhängigen Peer-Review-Verfahren unter den SCI-indexierten Zeitschriften höher als bei anderen. Dasselbe gilt für den Anteil, den SBPR-Verfahren an SCI-indexierten Zeitschriften haben. Gerade daran zeigt sich, dass Single Blind Peer Review für Zeitschriften aus dem naturwissenschaftlichen und medizinischen Bereich de facto das Standardverfahren ist.

## Alles „Open“?

Die weitgehende Verlagerung des Publikationsprozesses in die digitale Welt ermöglicht nicht nur die effizientere Unterstützung der bisher bereits etablierten Verfahren – insbesondere durch den Einsatz webbasierter und Email-gestützter Softwaresysteme im Rahmen der qualitativen Begutachtung. Viele neuartige Bewertungsverfahren sind darüber hinaus unter Nutzung elektronischer Kommunikationsstrukturen und des Online-Publizierens bzw. unter der Voraussetzung der schrankenlosen Verfügbarkeit der wissenschaftlichen Arbeiten überhaupt erst realisierbar. Dazu zählen vor allem Ansätze, die durch ein hohes Maß an

Offenheit, allgemeiner Partizipation und Interaktivität gekennzeichnet sind. An erster Stelle ist dabei das so genannte *Open Peer Commentary* (OPC) zu nennen, das grundsätzlich darauf beruht, dass einer vergleichsweise großen Anzahl von Wissenschaftlern die Möglichkeit eingeräumt wird, an der Bewertung wissenschaftlicher Arbeiten teilzuhaben. Dies geschieht im Gegensatz zu den meisten Ausprägungen herkömmlicher Peer-Review-Verfahren in der Regel ohne dezidierte Aufforderung seitens des Herausgebers. Stattdessen können ausgewiesene und in geeigneter Weise autorisierte Fachkollegen bzw. beliebige Rezipienten aus eigener Initiative heraus Kommentare zu einer Arbeit abgeben, die – je nach Aufbau des Verfahrens – den Status eines Manuskripts oder einer bereits formal publizierten Veröffentlichung besitzt.

Es ist zunächst durchaus erwartbar, dass sich sowohl nachgelagerte OPC-Verfahren als auch im Rahmen bzw. in zeitlicher Parallelität zu der formalen Qualitätssicherung stattfindende offene Prozesse für Open-Access-basierte Publikationsformen als konsequente Fortführung von Transparenz und Offenheit schnell durchsetzen und etablieren. Tatsächlich lässt sich zwar eine deutliche Tendenz hin zum freien Zugang zu wissenschaftlichen Veröffentlichungen beobachten, der sich insbesondere auch an der zunehmenden Verbreitung von Open-Access-Zeitschriften sowie hybriden Geschäftsmodellen manifestiert. Deren Akzeptanz innerhalb der einzelnen Wissenschaftsgemeinschaften hinsichtlich wissenschaftlicher Qualität sowie in Bezug auf das grundsätzliche Renommee steigt durchaus. Offene Kommentar- bzw. Begutachtungsformen werden dagegen bisher extrem schwach genutzt und finden seitens der angesprochenen Nutzerzielgruppe fast keine Resonanz, wie eine ebenfalls im beschriebenen Zusammenhang durchgeführte empirische Untersuchung ergab.

Die grundsätzlichen Probleme, die mit offenen Bewertungsverfahren verknüpft sind, lassen sich vor allem auf zwei Aspekte zurückführen – die Sicherstellung der Autorität der jeweiligen Teilnehmer sowie die Integration der Aktivitäten in diesem Zusammenhang in das



Anerkennungssystem der Wissenschaft insgesamt. Während das Vertrauen in die herkömmlichen Ansätze zur Qualitätssicherung und damit das Renommee, das eine Zeitschrift insgesamt genießt, wesentlich auf das Vertrauen in den Herausgeber und auf dessen anerkannte Autorität zurückgehen, ist bei den meisten OPC-Verfahren für Dritte schwer zu erkennen, inwieweit einem abgegebenen Kommentar oder gar einer lediglich auf einer linearen Skala erfolgten Bewertung tatsächlich eine fachliche Expertise zugrunde liegt. Selbst wenn Kommentare erst nach einer erfolgten Identitätsprüfung abgegeben werden können und damit eine Authentifizierung stattfindet, wird damit nicht ohne weiteres eine tatsächliche Autorität der jeweiligen Bewertungen erreicht. Auf der anderen Seite fließen Kommentare, die Wissenschaftler im Rahmen von OPC-Verfahren abgeben, in keiner Weise in das formale Anerkennungssystem des Wissenschaftsbetriebs ein. Dadurch besteht seitens der angesprochenen Zielgruppe wenig Anreiz, selbst aktiv zu werden und sich mit einer eigenen Einschätzung einerseits öffentlich zu exponieren und andererseits gegebenenfalls zur Verbesserung einer Arbeit und zum Erkenntnisfortschritt insgesamt beizutragen, ohne dafür entsprechende Meriten zu erlangen. Dieser zweite Aspekt – die fehlende formale Einbettung von Kommentarfunktionen – scheint der wesentliche Grund dafür zu sein, dass die bislang erfolgten Versuche, solche Verfahren zu etablieren, fast durchweg erfolglos geblieben sind.

Dies bedeutet allerdings nicht, dass die Transparenz von Begutachtungs- und Bewertungsverfahren und deren grundsätzliche Offenheit in Bezug auf potentielle Gutachter bzw. Kommentatoren prinzipiell zum Scheitern verurteilt sind. Beide Aspekte ergänzen den Open-Access-Gedanken hinsichtlich einer stärkeren Öffnung der Wissenschaft, deren Ergebnisse und deren zugrunde liegender Prozesse. Dass offene Verfahren durchaus funktionieren können, zeigen die Beispiele für Open Peer Review der medizinischen Zeitschriften der *BMC Series*, das durchweg offengelegte Begutachtungsverfahren bei *Atmospheric Chemistry and Physics (ACP)* und anderer durch dieses

Verfahren realisierter Zeitschriften sowie auch das bereits wesentlich ältere OPC-Verfahren, das bei der Zeitschrift *Behavioral and Brain Sciences (BBS)* angewendet wird. Allen Fällen ist neben einer hohen Transparenz hinsichtlich der Kommentare bzw. Gutachten gemein, dass sie stark formalisiert und in das Gesamtverfahren eingebunden sind. Wissenschaftler werden seitens des jeweiligen Herausgebers explizit eingeladen bzw. aufgefordert, sich an der Bewertung zu beteiligen, wodurch eine hohe Verbindlichkeit geschaffen wird. Bei BBS haben die Kommentare den Status einer eigenen kleinen Publikation. Bei den *BMC Series* und *ACP/D* agieren die Wissenschaftler als formal eingesetzte Gutachter mit dem üblichen Einfluss auf die eigentliche Publikationsentscheidung, die jedoch ihre Gutachten – und im Falle von BMC auch ihre Identitäten – offenlegen müssen.

Dass diese deutliche Erhöhung der Transparenz im Vergleich zu etablierten Peer-Review-Verfahren insbesondere auch durch die Gutachter akzeptiert wird und bei renommierten und hoch zitierten Zeitschriften möglich ist – bzw. zu renommierten und hoch zitierten Zeitschriften führt – zeigt, dass eine Öffnung der Prozesse im Rahmen der Qualitätssicherung durchaus praktikabel und zumindest in einigen Wissenschaftsbereichen mit der Wissenschaftskultur kompatibel zu sein scheint.

## Fazit

Peer-Review-Verfahren sind als Mittel der Qualitätssicherung für wissenschaftliche Publikationen nach wie vor das Mittel der Wahl und trotz vielfältiger, teils grundsätzlicher Kritik aus Sicht der meisten Wissenschaftler praktisch alternativlos (siehe [4]). Entgegen anders lautender Vermutungen und Überzeugungen finden sie auch bei Open-Access-Zeitschriften weite Verbreitung. Allerdings ist das, was im Einzelfall tatsächlich unter Peer Review verstanden wird, sehr vielfältiger Natur – es wäre daher auch unangemessen, von *dem einen* Peer-Review-Verfahren zu sprechen. Vielmehr handelt es sich um einen ganzen Strauß

von Verfahren mit sehr unterschiedlichen Ausprägungen. Bedeutsam dabei ist die Erkenntnis, dass die Beschaffenheit von Begutachtungsverfahren in besonderer Weise von dem Fachgebiet abhängig ist, in dem eine Zeitschrift verortet ist. Sie deutet darauf hin, dass in Bezug auf das wissenschaftliche Publizieren und hinsichtlich der Qualitätssicherung fachspezifische Unterschiede stärker in den Blick genommen werden müssen und die Diskussion – beispielsweise auch im Zusammenhang mit der Verbreitung von Open Access im Allgemeinen – nicht pauschalisiert geführt werden sollte.

## Literatur

- [1] OVER, A.; MAIWORM, F.; SCHELEWSKY, A.: *Publikationsstrategien im Wandel?: Ergebnisse einer Umfrage zum Rezeptions- und Publikationsverhalten unter besonderer Berücksichtigung von Open Access*. Weinheim : Wiley, 2005.
- [2] HARNAD, S.: *The invisible hand of peer review*. Exploit interactive (5), 2000.
- [3] MÜLLER, U.: *Peer-Review-Verfahren zur Qualitätssicherung von Open-Access-Zeitschriften : Systematische Klassifikation und empirische Untersuchung*. Dissertation, Humboldt-Universität zu Berlin. URL: <http://edoc.hu-berlin.de/dissertationen/mueller-uwe-thomas-2008-12-17/>, 2008.
- [4] WARE, M.: *Peer review in scholarly journals – Perspective of the scholarly community - an international study*. Bristol, UK, 2008.

# Die neue Q-Serie: eine Ergänzung des edoc-Angebotes

Susanne Dobratz | [dobratz@cms.hu-berlin.de](mailto:dobratz@cms.hu-berlin.de)

Karin Schröder | [schroeder@ub.hu-berlin.de](mailto:schroeder@ub.hu-berlin.de)

## Was ist die Q-Serie ?

Jährlich entstehen an der Humboldt-Universität einige Hundert Dissertationen. Viele Promovierende nutzen seit 1998 die Möglichkeit, der Publikationspflicht ihrer Dissertation durch die Veröffentlichung der elektronischen Version auf dem edoc-Server der Humboldt-Universität nachzukommen. Die Grundlage dafür bildet ein Beschluss des Akademischen Senats vom 14. Februar 1998, in dem die elektronische Publikation entsprechend den Bedingungen der Universitätsbibliothek möglich gemacht wird: *„Zusätzlich zu den in den einzelnen Promotionsordnungen genannten Möglichkeiten gilt auch die Ablieferung von vier vollständigen Exemplaren, die auf alterungsbeständigem, holz- und säurefreiem Papier ausgedruckt und dauerhaft haltbar gebunden sind, sowie einer elektronischen Version, deren Dateiformat und Datenträger mit der Universitätsbibliothek abzustimmen sind, als Erfüllung der Pflicht zur Veröffentlichung und Verbreitung der Dissertation“*, vgl. [1] und [2].

Seitdem wurden insgesamt 1571 Dissertationen auf dem edoc-Server veröffentlicht (Stand: 9. Februar 2009).

Viele Autorinnen und Autoren haben allerdings den Wunsch geäußert, auch noch eigene Druckexemplare zu besitzen (zusätzlich zu den immer noch notwendigen 4 gedruckten Papierpflichtexemplaren, die an die Universitätsbibliothek geliefert werden müssen). Sie möchten diese ihren Fachkollegen zur Verfügung stellen oder ggf. auch noch weiter verbreiten. Bisher konnten wir diesem Wunsch nur eingeschränkt nachkommen, indem wir weitere Exemplare der A4-Druckver-

sion der Dissertation mit einer nicht besonders schönen Schwarzpappenbroschur anfertigen ließen.

Um diese Lücke zu schließen und auch dem berechtigten Wunsch nach attraktiven Papierexemplaren nachzukommen, haben wir die Q-Serie ins Leben gerufen. Die Q-Serie kommt einer Verlagspublikation gleich, sie verbindet die Vorteile einer elektronischen Open-Access-Publikation mit denen einer Verlagspublikation. Die edoc-Betreiber kommen damit dem Bedarf nach, attraktiv gestaltete Druckexemplare zusätzlich zu den elektronischen Vertriebswegen auch über den normalen Buchhandel und dessen Internetvertriebsplattformen zu verbreiten.

Seit Juli 2008 gehört das Publizieren in der Q-Serie zu einem neuen Angebot des edoc-Servers. Dissertationen und Habilitationen können damit als so genannte Hybridpublikation erscheinen. Das bedeutet, dass neben der Open-Access-Veröffentlichung auf dem edoc-Server auch eine Printpublikation mit einem einheitlichen Layout und zusätzlichen Eigenschaften, wie etwa einer ISBN-Nummer, oder dem Vertrieb über Libri BOD, Buch.de sowie Amazon im Paket der Q-Serie enthalten ist. Die Publikation in der Q-Serie ist kostenpflichtig. Dabei bildet das digitale Exemplar immer das Original bzw. das Kernexemplar, von dem die Printversion abgeleitet ist. Dieses digitale Exemplar wird in jedem Fall kostenfrei und für alle verfügbar, OpenAccess-Ausgabe, angeboten. Für die Q-Serie steht ein eigener Bereich/ Zugang auf dem edoc-Server bereit. Bisher sind 5 Dissertationen von Angehörigen der Humboldt-Universität unter der Flagge der Q-Serie erschienen.

Seit Juli 2008 bietet der edoc-Server als zusätzliche neue Dienstleistung für Dissertationen die Publikation in der Q-Serie an. Die Q-Serie beinhaltet neben dem Open-Access-Publizieren auf dem edoc-Server die Möglichkeit, eine im Corporate Design der Humboldt-Universität gestaltete, gedruckte Fassung der Dissertation in den Buchhandelsvertrieb einzubringen und über digitale Vertriebsplattformen zu verbreiten. Eine Publikation in der Q-Serie ist für den Autor kostenpflichtig.

**Keywords?**



Abb. 1: Druckversion der Q-Serie

## Der edoc-Server

Der edoc-Server arbeitet mit folgenden Zielsetzungen:

- Veröffentlichungen von Wissenschaftlern der Humboldt-Universität schnell und kostengünstig zu ermöglichen
- das wissenschaftliche Leistungspotential der Universität nachhaltig nach außen sichtbar zu machen
- wissenschaftliche Veröffentlichungen im Open-Access-Verfahren der Öffentlichkeit zur Verfügung zu stellen
- auch solche Publikationen zu ermöglichen, die auf dem kommerziellen Verlagsmarkt nicht veröffentlicht werden können.

Mit der Q-Serie versuchen wir, die Verbreitung der Publikationen auch für klassische Vertriebswege, wie den Buchhandel, zu forcieren. Unsere Autoren wollen gelesen werden. Wir unterstützen sie dabei.

## Wer kann in der Q-Serie veröffentlichen?

Eine Publikation in der Q-Serie steht allen Promovierenden der Humboldt-Universität zu Berlin zur Publikation ihrer Dissertation offen, die ihr Promotionsverfahren erfolgreich absolviert haben.

Dabei können solche Dissertationen veröffentlicht werden, die das Prädikat „cum laude“ oder besser erreicht haben.

Für die Einreichung der Dateien gelten die gleichen Regelungen wie für das „normale“ Publizieren von Dissertationen auf dem edoc-Server. Denn auch eine Publikation in der Q-Serie ist nur im Zusammenhang mit einer elektronischen Publikation auf dem edoc-Server möglich.

Die Kosten für die Q-Serie werden individuell berechnet. Sobald das Dokument vorliegt, wird anhand der Seitenanzahl, der Auflage und der vorliegenden Formatierungsqualität ein Angebot erstellt. Dieses beinhaltet eine Grundgebühr sowie eine aufwandsabhängige Gebühr für den Druck und die Aufbereitung des Manuskriptes.

Der Autor bezahlt einmalig für die Erstellung einer Anfangsausgabe von 20 Exemplaren oder mehr (je nach Wunsch). Dabei liegt der Preis pro Exemplar deutlich unter dem Vertriebspreis des Buches.

## Eigenschaften des digitalen Angebotes

Den Kern der Q-Serie bildet das digitale Angebot. Dabei fließen die Publikationen der Q-Serie in den Dissertationsbereich auf dem edoc-Server ein, sie werden dort speziell gekennzeichnet.



Abb. 2: Icon; bei Auswahl wird ein Druckauftrag veranlasst. Erst einmal ein schneller Entwurf von mir!

Diese Publikationen können Open-Access von jedem jederzeit als durchsuchbares, ausdrucksfähiges PDF-Dokument heruntergeladen und weiter genutzt werden. Zusätzlich dazu werden sie auch als navigierbare Browservariante (HTML/XML) angeboten.

Jeder Autor vergibt für sein „Buch“ entsprechende Nutzungsrechte. Hierbei handelt es sich um Creative Commons-Lizenzen. Zur Auswahl stehen folgende Möglichkeiten:

Jede Publikation in der Q-Serie er-

1. Namensnennung	
Nichtkommerziell	
Keine Bearbeitung	
2. Namensnennung	
Nichtkommerziell	
Weitergabe unter gleichen Bedingungen	

hält eine URN in Form der DNB-Vorgaben als eindeutigen Identifikator des digitalen Exemplars. Zusätzlich dazu erhält das gedruckte Exemplar eine ISBN-Nummer und wird in den Buchhandelsvertrieb aufgenommen.

Zu den einzelnen Publikationen werden dokumentbezogene Zugriffsstatistiken angeboten. Die bibliografischen

Daten sind in verschiedenen Bibliografieformaten downloadbar (BibTex, Citavi, Reference Manager, Endnote, MARC21).

Sämtliche Publikationen der Q-Serie werden für das automatische Zählverfahren der VG Wort/METIS angemeldet.

Sie unterliegen der automatischen Meldung der Netzpublikation bei der Deutschen Nationalbibliothek in Frankfurt/Main und sind dort auch langfristig verfügbar. Gleichzeitig finden diese Publikationen Eingang in das Langzeitarchivierungs-Procédere des edoc-Servers der Humboldt-Universität zu Berlin.

schulzeitschriften wie der der „HUMBOLDT – Die Zeitung der Alma Mater Berolinensis“ vorgenommen werden und dadurch, dass Rezensionsexemplare an Fachzeitschriften (auf Anforderung) versendet werden.

Die Publikationen der Q-Serie sind über Google-Books auffindbar. Eine direkte Bestellmöglichkeit des Printexemplars über den edoc-Server ist gegeben.

Weitere Informationen erhalten Sie auf unseren Webseiten: [http://edoc.hu-berlin.de/formatierungsservice/e\\_autoren/formatierung.php](http://edoc.hu-berlin.de/formatierungsservice/e_autoren/formatierung.php)

Hier werden Sie betreut:

Karin Schröder, Susanne Dobratz  
formatierungsservice@hu-berlin.de  
Tel. 030/2093 70074 oder -70070

[http://edoc.hu-berlin.de/formatierungsservice\\_autoren/formatierung.php](http://edoc.hu-berlin.de/formatierungsservice_autoren/formatierung.php)

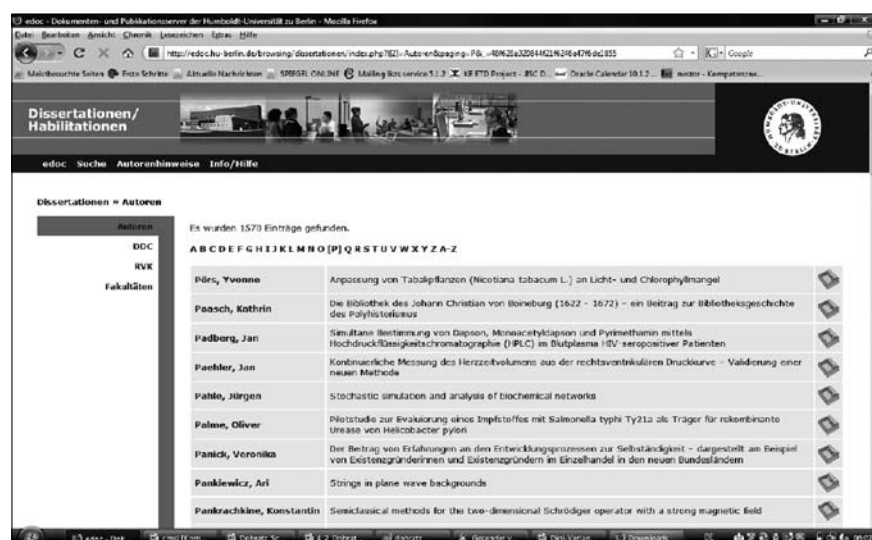


Abb. 2: Online-Darstellung der Q-Serie auf edoc (Dummybild, echte Version wird erst auf edoc erstellt)

## Q-Serie als Printauflage

Als Mindestauflage sind 20 Exemplare vorgesehen, davon gehen 4 Exemplare an die Universitätsbibliothek zur Erfüllung der Ablieferungspflicht. Alle anderen Exemplare erhält der Autor. Die Exemplare der Q-Serie werden im DIN-A5-Format, doppelseitig bedruckt und mit einem kaschierten Cover versehen.

Die Arbeitsgruppe Elektronisches Publizieren als Betreiberin des edoc-Servers meldet alle Titel an den Buchhandel (VLB) sowie an Verkaufsportale wie Amazon.

Sie sorgt für die Aufnahme in die Bibliothekskataloge der UB und darüber über die Verbreitung in Verbunddatenbanken etc. Die Marketingbemühungen der Autoren werden dadurch unterstützt, dass die Ankündigung neuer Publikationen der Q-Serie in Hoch-

## Literatur

- [1] Der Erste Vizepräsident: „Fachübergreifende Verfahrensregel über die Veröffentlichung von Promotionen“, in Amtliches Mitteilungsblatt der Humboldt-Universität zu Berlin, AMB 14/ 1998, <http://www.amb.hu-berlin.de/1998/14/1419980> (zuletzt besucht 9. Februar 2009)
- [2] siehe Text auf dem edoc-Server: [http://edoc.hu-berlin.de/e\\_autoren/as.php?arbeit=Dissertationen%20%C2%BB&index=index.php&nav=diss](http://edoc.hu-berlin.de/e_autoren/as.php?arbeit=Dissertationen%20%C2%BB&index=index.php&nav=diss) (zuletzt besucht am 07.02.2009)

# Der „Grüne Weg zu Open Access“ – Aktuelle Entwicklungen in Deutschland

Uwe Müller | u.mueller@cms.hu-berlin.de

Peter Schirmbacher | schirmbacher@cms.hu-berlin.de

*Das selbständige Einstellen von wissenschaftlichen Arbeiten ins Netz wird gemeinhin als der „Grüne Weg zu Open Access“ bezeichnet. Dabei gibt es eine Reihe von Besonderheiten, die in diesem Beitrag näher beleuchtet werden, der damit auch die Einstimmung für die nachfolgenden Artikel zur Nutzung, zum Betrieb und zur Vernetzung von Repositorien gibt. Der Beitrag bespricht die wichtigsten Spielarten des Grünen Weges – die so genannten Institutional Repositories und Fachrepositories und vergleicht sie hinsichtlich ihrer spezifischen Eigenschaften miteinander. Anschließend werden die unterschiedlichen Sichtweisen von Autoren und Rezipienten auf Open-Access-Repositories diskutiert, bevor Erfolgsfaktoren für Repositories und mögliche zukünftige Entwicklungen benannt werden. Der Beitrag lehnt sich an einen früheren Aufsatz der beiden Autoren im Heft 4-5/2007 der Zeitschrift für Bibliothekswesen und Bibliographie an [1], wurde jedoch entsprechend aktualisiert und ergänzt.*

*Green Road to Open Access, Self Archiving, Institutional Repository, Disciplinary Repository, persönliche Webseiten, Rezipienten, Autoren, Betreiber*

## Auf dem Grünen Weg

Mit *Open Access* wird der freie, unmittelbare und uneingeschränkte Zugang zu wissenschaftlichen Publikationen und Forschungsergebnissen in elektronischer Form bezeichnet, wie dies unter anderem in der *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* („Berliner Erklärung“) [2] aus dem Jahre 2003 beschrieben wird. Um das Ziel des freien Zugangs zu wissenschaftlichen Veröffentlichungen zu erreichen, besteht einerseits die Möglichkeit, traditionelle Publikationsformen in Open-Access-basierte Geschäftsmodelle zu überführen, sie aber in ihrer Organisationsform im Grundsatz unverändert zu lassen. Prominentestes Beispiel dafür sind die Open-Access-Zeitschriften, die sich von anderen elektronisch erscheinenden Journalen nur dadurch unterscheiden, dass die Einnahmen nicht aus Subskriptionsgebühren generiert werden, sondern beispielsweise durch Autorenbeiträge. Diesen Ansatz bezeichnet man auch als den *Goldenen Weg* zu Open Access.

Allerdings stellt diese reinste Form von Open Access [3] in den meisten Fällen nicht den ersten Schritt dar [4]. Nach einem alternativen Ansatz, dem so genannten *Grünen Weg*, der häufig auch als *Self Archiving* bezeichnet wird, werden wissenschaftliche Arbeiten unabhängig von traditionellen Publikationswegen öffentlich bereitgestellt. Dies geschieht entweder vor, nach oder anstelle der formalen Veröffentlichung, beispielsweise in wissenschaftlichen Zeitschriften. Das bedeutet, dass Autoren ihre gegebenenfalls bereits andernorts erschienenen oder eingereichten Publikationen auf

ihrer eigenen Webseite (*Individual Self Archiving*), dem Publikationsserver ihrer Universität oder wissenschaftlichen Einrichtung (*Institutional Self Archiving*) oder auf fachlich ausgerichteten Plattformen (*Central Self Archiving*) ablegen und damit neben der möglicherweise zugangsbeschränkten (Erst-)Veröffentlichung gewissermaßen eine frei verfügbare Kopie ihrer Publikation bereitstellen.

Dies geschieht meist auf freiwilliger Basis, zuweilen jedoch auch auf der Grundlage von Empfehlungen, Erklärungen oder Verpflichtungen von Forschungseinrichtungen bzw. Förderinstitutionen. Einen Überblick über institutionelle Richtlinien zu Open Access gibt die *Registry of Open Access Repository Material Archiving Policies* (ROARMAP)<sup>1</sup>, die weltweit etwa 110 Regelungen<sup>2</sup> auflistet. Dazu zählen vor allem Universitäten, die ihre Wissenschaftler mit entsprechenden Aufrufen für das Thema sensibilisieren wollen und sie zur Open-Access-Veröffentlichung ihrer Artikel auffordern. Zu einer bindenden Verpflichtung in Bezug auf die Veröffentlichungsform oder zu der so genannten *Anbietungspflicht*<sup>3</sup> [5] ist es an deutschen Universitäten nicht zuletzt mit Verweis auf die Wissenschaftsfreiheit bisher nicht gekommen. In der ROARMAP sind Open-Access-Leitlinien von insgesamt

1 Die Liste findet sich unter <http://www.eprints.org/openaccess/policy/signup/> [11.01.2009].

2 Etwas mehr als die Hälfte davon verpflichtet die Wissenschaftler zur frei zugänglichen Veröffentlichung ihrer Artikel.

3 Mit der Anbietungspflicht ist eine gesetzliche Regelung gemeint, die an öffentlichen Hochschulen beschäftigte Mitarbeiter verpflichten soll, ihre wissenschaftlichen Artikel zunächst der eigenen Hochschule zur Veröffentlichung anzubieten.



acht deutschen Institutionen vertreten. Neben der DFG haben sich bisher fünf Universitäten – unter ihnen die Humboldt-Universität zu Berlin – und ein An-Institut derartige Richtlinien gegeben. Die verbindlichsten Regelungen erließ im Juli 2008 die Fraunhofer-Gesellschaft. Sie verpflichtet ihre Mitarbeiter unter anderem dazu, an Verleger keine exklusiven Verwertungsrechte an ihren Arbeiten mehr zu vergeben<sup>4</sup> und macht somit eine Bereitstellung aller Publikationen im Sinne von Open Access aus rechtlicher Sicht grundsätzlich möglich.

## Viele Wege führen zum Ziel: Ausprägungen des Grünen Weges

Wie im letzten Abschnitt erwähnt, werden drei Arten des Self Archiving im Sinne des Grünen Weges unterschieden: Individual Self Archiving, Institutional Self Archiving und Central Self Archiving. Daraus ergeben sich entsprechend unterschiedliche Typen von Repositories, wobei der Begriff Repository zunächst sehr breit gefasst wird: Ein *Repository* (engl. für Depot, Lager, Speicher oder Repositorium) soll hier als ein System verstanden werden, mit dem eine Sammlung elektronischer Objekte verwaltet wird und nach außen online zur Verfügung gestellt werden kann. Dabei handelt es sich in der Regel vor allem um textorientierte Dokumente, was andersartige digitale Objekte jedoch nicht ausschließt.

Wesentliche Funktionen von Repositories sind somit das Einstellen elektronischer Dokumente (*Ingest*), deren Speicherung (*Storage*) sowie der Zugang dazu (*Access*). Selbstverständlich können diese Grundfunktionen sehr unterschiedlich realisiert und ausgeprägt sein und durch eine Vielzahl weiterer Dienste ergänzt werden. Den Zugang zu den Dokumenten bildet in der Regel eine Web-Oberfläche, die sich an menschliche Benutzer richtet und u. a. Links zu den elektronischen Dokumenten enthält. Viele Repositories gewähren jedoch Zugang (auch) über eine maschinelle Schnittstelle,

um die Daten übergreifenden Diensten – beispielsweise einer Metasuche – zur Verfügung zu stellen.

### Die persönliche Webseite

Im Sinne dieser Definition des Begriffs „Repository“ stellt eine Sammlung eines einzelnen Wissenschaftlers mit Veröffentlichungen, eigenen Vorträgen und dergleichen auf seiner persönlichen Webseite bereits ein Repository dar, das über die genannten Grundfunktionen verfügt. Je nachdem, wie regelmäßig und intensiv eine solche Sammlung gepflegt wird und nach welchen Kriterien sie zusammengestellt ist, spiegelt sie die jeweils aktuelle Gesamtheit der Veröffentlichungen eines Wissenschaftlers oder lediglich entsprechende Ausschnitte daraus wider.

Qualität und Aufbau bestehender Webseiten hängen vornehmlich von persönlichen Vorlieben, verfügbaren personellen Kapazitäten und vorhandenem Know-how ab. Abgesehen davon ist der Gegenstandsbereich des Self Archiving sehr verschieden: Während viele persönliche Sammlungen lediglich Veröffentlichungen enthalten, die einen Peer-Review-Prozess durchlaufen haben und bereits publiziert sind, finden sich andernorts auch Vorträge, interne Berichte, Zeitungsinterviews und Ähnliches wieder.

Auch wenn das Individual Self Archiving im Rahmen der Diskussionen zu Open Access der Vollständigkeit wegen meist mit aufgeführt wird, werden persönliche Webseiten mit den elektronischen Versionen eigener Veröffentlichungen eher selten als Repository im engeren Sinne eingestuft. Denn wesentliche Anforderungen an das wissenschaftliche Publizieren – beispielsweise die Zitierbarkeit, die zuverlässige und dauerhafte Verfügbarkeit und die Wahrung von Authentizität und Integrität der Veröffentlichungen – sind in den meisten Fällen nicht erfüllt oder zumindest nicht garantiert und verbindlich festgeschrieben.

## Institutional Repositories

Nach den aktuellen Erhebungen der Deutschen Initiative für Netzwerkinformation (DINI)<sup>5</sup> verfügen in Deutschland nahezu alle Universitäten und eine Vielzahl von Fachhochschulen und außer-universitären Forschungseinrichtungen über ein so genanntes Institutional Repository – also einen Dokumentenserver, auf dem wissenschaftliche Publikationen der jeweiligen Einrichtung in elektronischer Form veröffentlicht werden.<sup>6</sup> Die Gesamtzahl beläuft sich derzeit auf etwa 140. Die meisten Repositories basieren auf einer der in Deutschland entwickelten Softwarelösungen OPUS oder MyCoRe. Darüber hinaus sind einige Installationen von eprints, DSpace und Fedora sowie Eigenentwicklungen im Einsatz.

Verantwortliche Betreiber dieser Institutional Repositories sind in der Regel die Universitätsbibliotheken, häufig in Verbindung mit dem Rechenzentrum. Für diese Einrichtungen stellt sich ein solcher Dokumentations- und Publikationsservice vielfach als eine zusätzliche Aufgabe der „digitalen Zeit“ dar, die von engagierten Kolleginnen und Kollegen aufgegriffen wurde und vielfach als Chance und Herausforderung begriffen wird.

Der Vorteil dieser Repositories besteht darin, dass sie in finanzieller und organisatorischer Hinsicht in der Regel auf soliden Füßen stehen. Der Dokumenten- und Publikationsservice wird in den meisten wissenschaftlichen Einrichtungen Deutschlands inzwischen als Regelangebot betrachtet.

Im internationalen Kontext können sich die deutschen Aktivitäten durchaus messen. Im Directory of Open Access Repositories (OpenDOAR), dem internationalen Verzeichnis aller Institutional Repositories<sup>7</sup>, werden derzeit 130 deutsche

<sup>5</sup> siehe <http://www.dini.de/wiss-publizieren/repository> [11.01.2009]

<sup>6</sup> Das muss nicht bedeuten, dass jede Einrichtung tatsächlich einen entsprechenden Server selbst betreibt. Gemeint ist hier vielmehr, dass ein entsprechendes Angebot zur elektronischen Veröffentlichung besteht, das beispielsweise auch gemäß einem Hosting-Ansatz durch einen externen Dienstleister realisiert werden kann.

<sup>7</sup> siehe <http://www.opendoar.org/> [11.01.2009]

<sup>4</sup> „Employees [...] are expressly required to demand that publishing houses grant them the right to further use their own works.“ [6]

Repositories ausgewiesen; damit steht Deutschland hinter den USA und Großbritannien mit 325 bzw. 143 registrierten Repositories weltweit an dritter Stelle.

Die weite Verbreitung von Institutional Repositories in der deutschen Hochschullandschaft darf allerdings nicht darüber hinweg täuschen, dass Anzahl und Häufigkeit der Nutzung dieser Publikationsmöglichkeit sehr unterschiedlich ausgeprägt und – verglichen mit dem tatsächlichen Aufkommen wissenschaftlicher Veröffentlichungen an den jeweiligen Einrichtungen – insgesamt noch nicht sehr entwickelt sind. Zu den am meisten auftretenden Dokumenttypen zählen nach wie vor Qualifikationsarbeiten (etwa Dissertationsschriften und Studienabschlussarbeiten) und Schriftenreihen, deren Herausgeberschaft bei den betreffenden wissenschaftlichen Institutionen angesiedelt ist.

Dagegen sind Pre- und vor allem Postprints, die den eigentlichen Gegenstand des Grünen Weges zu Open Access darstellen, in den einzelnen Institutional Repositories bislang kaum vorhanden. Das liegt zum einen an einem geringen Bewusstsein für das Thema sowohl seitens der Hochschulleitungen als auch bei den Wissenschaftlern. Darauf weisen unter anderem eine Umfrage der DFG unter Wissenschaftlern aus dem Jahr 2005 [7] als auch die Tatsache, dass nach der bereits erwähnten ROARMAP lediglich fünf deutsche Universitäten eine Open-Access-Erklärung verabschiedet haben, hin. Doch auch in diesen Fällen bleiben die entsprechenden Bereiche der Dokumentenserver bislang spärlich gefüllt.

Zum anderen führen oftmals ungeklärte rechtliche Fragen und der – bei zusätzlichem Aufwand – fehlende Anreiz, Veröffentlichungen auf dem Institutional Repository abzulegen, dazu, dass Autoren diesen Schritt bisher noch eher selten gehen. Andererseits dürfte es sich dabei nicht zuletzt um ein Vermittlungsproblem handeln. Denn dass Open-Access-Publikationen beispielsweise eine höhere Zitierhäufigkeit aufweisen [8], [9] und der Verzicht auf Open Access damit zum so genannten impact loss [10] führt, wurde inzwischen mehrfach statistisch nachgewiesen.

Selbst die Max-Planck-Gesellschaft, die als Erstunterzeichnerin und Mitinitiatorin der Berliner Erklärung diejenige außeruniversitäre Forschungseinrichtung ist, die in Bezug auf Open Access bislang die meisten Anstrengungen unternommen hat, stellt nur einen Bruchteil ihrer Publikationen im Volltext frei zur Verfügung: Lediglich für etwa 7,5 % der mehr als 120.000 Publikationen, die auf dem eDoc-Server der MPG nachgewiesen sind, ist auch der elektronische Volltext frei verfügbar.

### Fachspezifische Repositories

Deutlich anders verhält es sich beim so genannten Central Self Archiving, das durch den Betrieb fachspezifischer Repositories realisiert wird. Auf derartigen Servern werden im Gegensatz zu den anderen Formen des Self Archiving und unabhängig von der institutionellen Zugehörigkeit der Autoren wissenschaftliche Arbeiten eines bestimmten Fachgebietes oder einer Fächergruppe abgelegt.

Der Wegbereiter und weltweit bekannteste Vertreter für diese Form von Open Access ist der ursprünglich in Los Alamos entstandene und heute an der Cornell University beheimatete Preprint-Server arXiv.org. Dort werden bereits seit Beginn der 1990er Jahre die elektronischen Fassungen wissenschaftlicher Veröffentlichungen im Bereich der Hochenergiephysik abgelegt. Heute stellt der arXiv-Server mehr als 500.000 Volltexte aus den Fachgebieten Physik, Mathematik, Informatik und Biologie frei zugänglich bereit. Es gehört zum „guten Ton“ in der Community dieser Fachgebiete, geplante eigene Veröffentlichungen als Preprint – also noch vor Abschluss eines möglichen Begutachtungsverfahrens – dort abzuliegen und damit der wissenschaftlichen Öffentlichkeit zur Verfügung und zur Diskussion zu stellen.

Innerhalb Deutschlands haben sich fachspezifische Repositories vor allem in Form der Virtuellen Fachbibliotheken etabliert, die in Anlehnung an die ebenfalls auf nationaler Ebene organisierten Bibliotheken für Sondersammelgebiete (SSG-Bibliotheken) entstanden sind. Die Virtuellen Fachbibliotheken bilden fachlich ausgerichtete Webportale, auf denen

Informationen und wissenschaftliches Material zu je einem Wissenschaftsgebiet zusammengeführt sind und aufgefunden werden können.

Repositories, auf denen elektronische Volltexte abgelegt und vorgehalten werden, bilden in diesen Fachportalen allerdings nur eine Informationsquelle unter vielen. Sie werden durch die Anbindung bibliografischer Kataloge und anderer Datenbanken, verzeichnete und indexierte Internetquellen und vielfach durch zusätzliche Dienste und Kommunikationsmittel ergänzt – wie etwa Annotationen und Rezensionen, Diskussionsforen oder Blogs.

Die Informationen, die die derzeit mehr als 30 Virtuellen Fachbibliotheken in Deutschland bereitstellen, und weitere Fachinformationsdienste werden unter dem Dach des vascoda-Portals<sup>8</sup> zusammengeführt. Vascoda, inzwischen ein eingetragener Verein, bündelt das Angebot für die wissenschaftliche Informationsversorgung vorrangig im deutschsprachigen Raum und ermöglicht dabei unter anderem die fachübergreifende Recherche. Dies ist besonders für interdisziplinäre Forschungen und Forschungsfelder, die sich an der Nahtstelle formaler Wissenschaftsgebiete bewegen, von großer Bedeutung.

Insgesamt lässt sich feststellen, dass die Virtuellen Fachbibliotheken innerhalb der jeweiligen Communities gut positioniert und als Anlaufstelle für wissenschaftliche Fachinformation relativ gut angenommen sind. Dennoch bestehen zumindest zwei Problemkreise, die für die in Deutschland existierenden fachspezifischen Repositories charakteristisch sind und eine Ursache dafür bilden, dass diese Entwicklung auf internationaler Ebene bislang eine weit geringere Verbreitung gefunden hat.

- Virtuelle Fachbibliotheken und als einer deren Bestandteile damit auch fachspezifische Repositories sind in Deutschland fast durchgängig unter Zuhilfenahme einer Anschubfinanzierung durch die Deutsche Forschungsgemeinschaft entstanden. Damit konnten die Anfangshürden zunächst erfolgreich genommen werden. Spätestens

<sup>8</sup> siehe <http://www.vascoda.de/> [11.01.2009]

nach Ablauf des Förderungszeitraums ergibt sich jedoch die Problematik der Verfestigung der entwickelten Plattformen.

- Das zweite Problem fachspezifischer Repositories, das auf den ersten Blick als rein wissenschaftspolitische Fragestellung gesehen wird, ist die fehlende fachliche Gesamtabdeckung sowie die fachliche Abgrenzung zwischen einzelnen Virtuellen Fachbibliotheken. Unter den ersten Bereich fallen solche Fachgebiete, für die es bisher keine entsprechenden Fachrepositories bzw. Virtuellen Fachbibliotheken gibt. Dieses Problem lässt sich letztlich nur durch den Aufbau weiterer Virtueller Fachbibliotheken oder durch die fachliche Verbreiterung bestehender Angebote angehen. Die Problematik der fachlichen Abgrenzung spiegelt die Kehrseite der Vorzüge fachspezifischer Repositories wider.

### Institution vs. Community

In den vorangegangenen Abschnitten wurden fachspezifische und Institutional Repositories vorgestellt. Es ist deutlich geworden, dass beide Ansätze ihre Vor- und Nachteile mit sich bringen. Während Institutional Repositories Spiegelbilder ihrer jeweiligen Forschungseinrichtungen und somit in der Regel fachlich sehr heterogen zusammengesetzt sind und nur einen Bruchteil der für ein Fach relevanten Publikationen vorhalten, zeichnen sich fachspezifische Repositories durch ihre inhaltliche Kohärenz und einen Anspruch auf relative Vollständigkeit aus. Dagegen haben Institutional Repositories in Bezug auf Nachhaltigkeit im Sinne des Fortbestandes sowie in Bezug auf die potentielle fachliche Gesamtabdeckung klare Vorzüge.

Sicherlich wird es nicht das eine Portal geben, über das Wissenschaftler in Zukunft alle aus ihrer Sicht jeweils relevanten Informationen und Publikationen beziehen können und werden. Die beiden Ansätze haben dennoch ihre je eigene Berechtigung und Gültigkeit und können, um die jeweiligen Vorteile auszunutzen und die Nachteile zu minimieren, ohne weiteres miteinander kombiniert werden. Ein sinnvoller Weg bestünde beispielsweise darin, dass Wissenschaftler

ihre Publikationen primär auf den für sie zuständigen Institutional Repositories ablegen, von wo aus sie unmittelbar und automatisch durch eine entsprechende fachliche Einordnung an die jeweiligen fachlichen Repositories weitergeleitet werden. Das setzt einheitliche Schnittstellen und einen Grundkonsens über eine grobe Fachklassifikation voraus.

In jedem Fall dürfen die Wissenschaftler und deren unterschiedliche Bedürfnisse – seien sie nun in der Rolle eines Autors oder eines Lesers – nicht aus dem Blick geraten. Denn nur wenn sich die Publikationslandschaft daran ausrichtet, wird sich der vielfach auf freien und individuellen Entscheidungen beruhende Publikationsprozess in die Richtung, in der ein Großteil der Publikationen elektronisch frei zugänglich ist, ausrichten. Im folgenden Abschnitt wird es darum gehen, wie sich die Repositories aus dem Blickwinkel der Beteiligten darstellen.

### Gegenwärtiger Stellenwert von Repositories

Bei aller durch die bisherigen Darlegungen verdeutlichten Wertschätzung der Repositories an den wissenschaftlichen Einrichtungen in Deutschland ist es im Hinblick auf die Fortentwicklung dieser Veröffentlichungsform wichtig, eine kritische Bestandsaufnahme vorzunehmen. Deren Ziel ist es, daraus Schlussfolgerungen für weitere Entwicklungsaufgaben ziehen zu können. Diese kritische Bestandsaufnahme soll Probleme derzeit bestehender Repositories aus drei unterschiedlichen Blickwinkeln aufzeigen und untersuchen:

1. aus Sicht von Rezipienten,
2. aus Sicht von Autoren und
3. aus Sicht von wissenschaftlichen Einrichtungen und Betreibern.

#### Repositories aus Sicht von Rezipienten

Auf der Suche nach wissenschaftlichen Fachinformationen bzw. bei der Literaturrecherche im Internet wird der durchschnittliche Nutzer in der Regel lediglich auf die fachspezifischen und gegebenenfalls auf institutionelle Repositories zurückgreifen. Nur in den seltensten Fällen

werden persönliche Webseiten mit Publikationslisten einzelner Wissenschaftler oder Arbeitsgruppen unmittelbar in die Recherche mit einbezogen.

Selbst die meisten Institutional Repositories werden nur in Ausnahmefällen Ausgangspunkt für eine ernsthafte Recherche nach wissenschaftlichen Inhalten sein. Der Grund dafür liegt auf der Hand: Die derzeit existenten Institutional Repositories umfassen jeweils nur wenige hundert oder ein paar tausend Dokumente, die zudem die fachliche Diversität der jeweiligen Forschungseinrichtung abbilden. Sie enthalten somit zumindest im Falle von Universitäten aus der Sicht eines einzelnen Wissenschaftlers verhältnismäßig wenig relevante Publikationen.

Die fehlende Abdeckung innerhalb eines Fachgebiets ist zudem eine immanente Eigenschaft des Ansatzes, den Institutional Repositories verfolgen: Sie veröffentlichen wissenschaftliche Arbeiten, die an der eigenen Einrichtung entstanden sind. Insofern sind Institutional Repositories zwar ein wichtiger Bestandteil in einer Open-Access-Publikationsinfrastruktur. Ihre Benutzerinterfaces werden aber nur selten Ausgangspunkt für Recherchen nach wissenschaftlicher Literatur sein. Umso wichtiger ist es, dass die Metadaten und damit auch die Dokumente selbst über definierte Schnittstellen übergreifenden Recherchediensten zur Verfügung gestellt werden.

Darüber hinaus ist ein wichtiger Kritikpunkt an vielen bestehenden Repositories, dass nur wenige Möglichkeiten der Einordnung einzelner Publikationen in einen Gesamtzusammenhang eröffnet werden. Oft fehlen aussagekräftige Nutzungsstatistiken oder eine strukturierte Aufbereitung bibliografischer Angaben – beispielsweise zum Zwecke einer Zitationsanalyse – die eine solche Einordnung erleichtern würden. Als großes Handicap vor allem der Institutional Repositories wird darüber hinaus eine fehlende qualitative Einordnung der Veröffentlichungen gesehen: Während bei Artikeln, die in einer wissenschaftlichen Zeitschrift erschienen sind, zumindest klar ist, ob und in welcher Form ein Peer-Review-Prozess durchlaufen wurde, fehlt eine Aussage über eine fachliche Begutachtung der Publikationen in einem In-

stitutional Repository meist vollständig. Das heißt zwar nicht, dass es sich dabei ausschließlich oder überwiegend um nicht begutachtete oder qualitativ minderwertige Dokumente handelt. Allein die fehlende Unterscheidungsmöglichkeit nach derartigen Qualitätskriterien macht eine Einbeziehung von Institutional Repositories in Rechercheinstrumente nicht gerade attraktiv.

Die meisten der genannten Probleme treffen die fachspezifischen Repositories weniger hart. Sie sind inhaltlich, aber auch methodisch auf ihre jeweilige Zielgruppe zugeschnitten und darauf ausgerichtet, zu einem Fachgebiet alle wesentlichen verfügbaren Informationsquellen zusammenzuführen. Darüber hinaus besitzen Virtuelle Fachbibliotheken vielfach Fachredaktionen, die einen Peer-Review-Prozess organisieren oder zumindest qualitative Bewertungen der einzelnen Dokumente und Quellen vornehmen können.

Häufig wird die Authentizität und Integrität der Dokumente jedoch skeptisch betrachtet und die Beständigkeit des Angebotes in Zweifel gezogen. Obwohl die Vorwürfe sachlich nicht unbedingt gerechtfertigt sind, ist es sehr sinnvoll, entsprechende technische Maßnahmen umzusetzen, die die Authentizität und Integrität von Dokumenten beweiskräftig sicherstellen. Entsprechende Vorgaben sind beispielsweise auch in den Kriterienkatalog des DINI-Zertifikats eingeflossen.

Bedauerlicherweise befindet sich die Mehrzahl der deutschen Repositories noch in einer ersten Entwicklungsphase, sodass die Zahl der Mehrwert-Dienste, wie z. B. ein Print-on-Demand-Angebot oder eine qualifizierte Suche stark begrenzt ist.

### Repositories aus Sicht von Autoren

Die Motive, aus denen heraus Autoren wissenschaftliche Arbeiten veröffentlichen, sind vielfältig. Die formalisierte wissenschaftliche Kommunikation mit Fachkollegen, aber auch die allgemeine Reputation und Anerkennung innerhalb der eigenen Community zählen jedoch zu den wesentlichen Gründen. In diesem Zusammenhang stehen die möglichst weit reichende Verbreitung der eigenen

Publikationen und die damit verbundene Rezeption und Zitation als Richtmaß im Vordergrund.

Betrachtet man die unterschiedlichen Möglichkeiten, die sich als Publikationsform für wissenschaftliche Veröffentlichungen anbieten, kommt man nicht umhin festzustellen, dass Institutional Repositories dabei in der Regel nicht die erste Wahl darstellen. Unter den unterschiedlichen Arten von Repositories erfüllen die fachspezifischen Repositories die genannten Anforderungen an die wissenschaftliche Kommunikation aus Sicht von Autoren am ehesten.

Sofern derartige Repositories in der jeweiligen Community gut etabliert sind – wie es beispielsweise im Falle von arXiv.org zutrifft – werden fachspezifische Repositories durchaus als Veröffentlichungsort gewählt. Die Veröffentlichung wissenschaftlicher Beiträge auf Institutional Repositories trifft sich dagegen in der Regel nicht mit der Interessenlage der Wissenschaftler, das heißt, es mangelt an einem entsprechenden Anreiz dafür.

Dies bedeutet nicht, dass Institutional Repositories zum Scheitern verurteilt sind. Sie sind jedoch eher von einem Sammlungscharakter geprägt und können dabei das wissenschaftliche Profil einer Einrichtung widerspiegeln. Im Idealfall bilden Institutional Repositories ein wichtiges Aushängeschild dieser Institution und die Datengrundlage für eine institutsübergreifende – auch fachlich spezifische – Informationsversorgung. Denn auch Virtuelle Fachbibliotheken können auf den jeweils relevanten Bestandteil zurückgreifen, der in Institutional Repositories vorgehalten wird. Damit sich Autoren auf freiwilliger Basis für eine – ggf. zusätzliche – Veröffentlichung ihrer wissenschaftlichen Arbeiten auf ihrem Institutional Repository entscheiden, muss ihnen jedoch deutlicher gemacht werden, inwiefern dies die eigene Interessenlage trifft.

Ohne jeden Zweifel ist anzuerkennen, dass die mit Abstand wichtigste Publikationsform in den meisten Wissenschaftsbereichen die traditionelle Zeitschriftenpublikation darstellt. Dabei werden nach wie vor die meisten Zeitschriften mithilfe von Verlagen und auf der Basis her-

kömmlicher Geschäftsmodelle realisiert. Open-Access-Zeitschriften sind dagegen prozentual noch in der Minderzahl und entwickeln sich erst sehr langsam.

Institutional Repositories kommen für die meisten Autoren in erster Linie als Möglichkeit für eine Zweitpublikation in Betracht. So lange in dieser ergänzenden Veröffentlichung kein Mehrwert erkannt wird, werden Wissenschaftler den zusätzlichen Aufwand scheuen. Gerade deshalb besteht eine wichtige Aufgabe für die Betreiber von Repositories darin, möglichst viele unterstützende Werkzeuge für Autoren anzubieten, um ihnen den Zugang zu den Repositories so leicht wie möglich zu gestalten.

### Erfolgsfaktoren für Repositories

Einerseits ist es durchaus erfreulich, dass in vielen wissenschaftlichen Einrichtungen in Deutschland der Stellenwert des Betriebes fachspezifischer und institutioneller Repositories im Rahmen der traditionellen Aufgaben von Infrastruktureinrichtungen immer mehr steigt. Dennoch ist in vielen Fällen der Übergang dieses zusätzlichen Aufgabenfeldes vom Projektstatus in den dauerhaft abgesicherten Regelbetrieb noch nicht bewältigt.

Aufbau und Betrieb von Dokumenten- und Publikationsservices stellen an die Universitätsbibliotheken und – wenn sie mit einbezogen werden – an die universitären Rechenzentren erhöhte Anforderungen. Dazu gehören neben Basiswissen aus der bibliothekswissenschaftlichen Sicht in vielen Fällen auch ausgeprägte Kenntnisse der Informatik und des Betriebs von Servern und dergleichen mehr. Insgesamt ist festzustellen, dass es in Deutschland noch zu wenige Kolleginnen und Kollegen gibt, die über entsprechende Qualifikationen für diese Art von Aufgaben verfügen.

Während die Mehrzahl der fachspezifischen Repositories durch Institute oder Fachbereiche von Universitäten betrieben wird und damit auf ein fachwissenschaftliches Know-how zurückgreifen können, lässt sich Vergleichbares für Institutional Repositories nicht sagen. Die Betreiber dieser Publikationsserver sind keine Fachwissenschaftler und verfügen



somit nur über einen sehr beschränkten Einfluss auf die Qualität des Inhalts und die Autoren selbst. Aus diesem Grunde orientieren viele Betreiber in erster Linie auf die Quantität als Maßstab für den Erfolg eines Institutional Repositories. Für sie stehen eher die Gesichtspunkte des Sammelns als die des Anbietens elektronischer Dokumente im Vordergrund.

Problematisch daran ist vor allem, dass die Entwicklung hochwertiger Services dabei oft in den Hintergrund tritt und nicht zu den vorrangigen Zielen des Aufbaus von Institutional Repositories gezählt wird.

Eine neuere Studie des Centers for Informatics Research in Science and Scholarship der University of Illinois at Urbana-Champaign [11] unterstützt diese Aussagen. Deutlich mehr als in Deutschland üblich, wird in den USA jedoch auch die Rolle der Professorinnen und Professoren im Zusammenwirken mit der Bibliothek gesehen. Sie sind die Vorbilder (und vor allem begreifen sie sich auch als solche) im wissenschaftlichen Kommunikationsverhalten. Sie prägen die jeweilige Community und natürlich zuvörderst das unmittelbare wissenschaftliche Umfeld.

## Zusammenfassung und Ausblick

Der „Grüne Weg zu Open Access“ ist in Deutschland durchaus sichtbar. In nahezu allen wissenschaftlichen Einrichtungen gibt es Bestrebungen, durch den Aufbau von Servern und einem entsprechenden Service ein Angebot für Autoren und Leser vorzuhalten. In diesem Sinne braucht Deutschland den internationalen Vergleich in keiner Weise zu scheuen. Unterzieht man die vorhandenen Repositories jedoch einer kritischen Analyse, so muss man erhebliche Defizite konstatieren, die sich insbesondere in der mangelnden Sichtbarkeit und in den teilweise fehlenden Mehrwertdiensten festmachen lassen.

Die Sichtbarkeit kann dadurch am besten erhöht werden, dass der Dokumentenraum erweitert wird und die einzelnen Repositories zu einem Netzwerk mit gemeinsamen Diensten verbunden werden. Seit Ende 2007 läuft dazu ein

durch die DFG gefördertes Projekt mit dem Titel „Aufbau eines Netzwerkes für zertifizierte Open-Access-Repositories“. Unter der Koordination von DINI e.V. verfolgen die Projektpartner aus der Humboldt-Universität zu Berlin, der Universität Osnabrück und der Niedersächsischen Staats- und Universitätsbibliothek Göttingen das Ziel, ein Netzwerk zertifizierter Repositories aufzubauen, in dem Recherchedienste – wie eine übergreifende Volltextsuche und ein klassifizierter Browsingzugang – sowie Mehrwertdienste – wie eine Zitationsanalyse und statistische Auswertungen – gemeinsam angeboten werden (siehe [12]).

Dieses Netzwerk bildet darüber hinaus den deutschen Knoten für ein gesamteuropäisches Vorhaben im Rahmen des EU-geförderten Projektes DRIVER (Digital Repository Infrastructure Vision for European Research). Außerdem soll mit Unterstützung durch das Projekt die Zahl der zertifizierten Repositories in Deutschland erhöht werden, um damit einen weiteren Beitrag zur Verbesserung der Sichtbarkeit zu leisten.

Sowohl der bisherige Umfang der ROARMAP als auch der Inhalt der existierenden Repositories von Institutionen lassen erkennen, dass der Grüne Weg zu Open Access nach wie vor ein Schattendasein führt und von einem durchschlagenden Erfolg wohl noch nicht gesprochen werden kann. Er stellt nach Ansicht vieler Experten ohnedies nur eine Zwischenstation auf dem Weg zu Open-Access-Zeitschriften dar [4].

## Literatur

- [1] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html> [11.01.2009].
- [2] FRAUNHOFER-GESELLSCHAFT: *The Fraunhofer-Gesellschaft's Open Access Policy*. 2008. [http://www.fraunhofer.de/fhg/Images/OpenAccessPolicy\\_tcm6-101804.pdf](http://www.fraunhofer.de/fhg/Images/OpenAccessPolicy_tcm6-101804.pdf) [11.01.2009]
- [3] GOODMAN, D.: *The Criteria for Open Access*. *Serials Review* 30 (4), 2004, S. 258–270.

- [4] GUÉDON, J.-C.: *The "Green" and "Gold" Roads to Open Access*. *Serials Review* 30 (4), 2004, S. 315–328.
- [5] HAJJEM, C.; HARNAD, S.; GINGRAS, Y.: *Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact*. *IEEE Data Engineering Bulletin* 28 (4), 2005, S. 39–47.
- [6] HANSEN, G. (2005): *Urheberrecht für Wissenschaftler – Risiken und Chancen der Urheberrechtsreform für das wissenschaftliche Publizieren*. In Stempfhuber, M. (Hrsg.) *In die Zukunft Publizieren*. II. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland. Bonn: Informationszentrum Sozialwissenschaften, 2005, S. 9–19.
- [7] HARNAD, S.; BRODY, T.; VALLIERES, F.; CARR, L.; HITCHCOCK, S.; GINGRAS, Y.; OPPENHEIM, C.; STAMERJOHANN, H.; HILF, E.: *The Access/Impact Problem and the Green and Gold Roads to Open Access*. *Serials Review* 30 (4), 2004, S. 310–314.
- [8] KURTZ, K.; BRODY, T.: *The impact loss to authors and research*. In Jacobs, N. (Hrsg.) *Open Access: Key strategic, technical and economic aspects*. Oxford: Chandos Publishing, 2006, S. 55–62.
- [9] MALITZ, R. (2009): *Open Access – Verfügbar ist nicht präsent*. *cms-journal* 32, 2009, S. 70–73.
- [10] MÜLLER, U.; SCHIRMBACHER, P.: *Der „Grüne Weg zu Open Access“ in Deutschland*. *Zeitschrift für Bibliothekswesen und Bibliografie* 54 (4-5), 2007, S. 199–206.
- [11] OVER, A.; MAIWORM, F.; SCHELEWSKY, A.: *Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access*. Studie der Deutschen Forschungsgemeinschaft. Weinheim: Wiley-VCH, 2005.
- [12] PALMER, C. C.; TEFFEAU L. C.; NEWTON M. P.: *Identifying Factors of Success in CIC Institutional Repository Development – Final Report*. Center for Informatics Research in Science and Scholarship, University of Illinois at Urbana-Champaign, 2008.



# Wie entsteht ein Institutional Repository? – Eine systematische Hinführung in acht Schritten

Susanne Dobratz | [dobratz@cms.hu-berlin.de](mailto:dobratz@cms.hu-berlin.de)  
Uwe Müller | [u.mueller@cms.hu-berlin.de](mailto:u.mueller@cms.hu-berlin.de)

## Einleitung

Ein digitales Repository gehört heute zur Standardausstattung nahezu jeder wissenschaftlichen Einrichtung. Fast alle Universitäten und viele Fachhochschulen, Forschungseinrichtungen und Informationseinrichtungen bieten Publikationen – zumeist solche eigener Mitarbeiter bzw. Studierender – in elektronischer Form an. Aufgrund der Vielzahl und der Verschiedenheit von Repositories, die nicht immer durch die jeweiligen Einrichtungsbibliotheken betrieben werden und von denen es an vielen Einrichtungen gleich mehrere gibt, und wegen der hohen Dynamik in diesem Bereich ist es kaum möglich, einen umfassenden Überblick über den aktuellen Stand in Deutschland zu erlangen. Das *Registry of Open Access Repositories*<sup>1</sup> (ROAR), das von der University of Southampton betrieben wird, listet derzeit 97 deutsche Dokumentenserver auf. Eine von DINI gepflegte Übersicht<sup>2</sup> enthält derzeit sogar 138 Einträge, von denen insgesamt 22 das DINI-Zertifikat tragen, sechs davon dessen aktuelle Version aus dem Jahr 2007. Größere Einrichtungen wie Universitäten verteilen ihre elektronischen Publikationen zuweilen auf mehrere Repositories – beispielsweise einen Medienserver zur Verwaltung multimedialer Lehr- und Lernmaterialien, einen Hochschulschriftenserver zur Veröffentlichung universitärer Qualifikationsarbeiten und eine Publikationsplattform für Pre- und Postprints wissenschaftlicher Veröffentlichungen.

Belastbare exakte Zahlen sind auch deshalb schwer zu ermitteln, weil viele Repositories Teil eines größeren Angebots, z. B. eines Fachportals und in den entsprechenden Verzeichnissen nicht als separater Publikationsserver gemeldet sind.

## Was ist ein digitales Repository?

Ein *Repository* (engl. für Depot, Lager, Speicher oder Repositorium) wird als System verstanden, mit dem eine Sammlung elektronischer Objekte verwaltet und nach außen online zur Verfügung gestellt werden kann. Dabei handelt es sich in der Regel vor allem um textorientierte Dokumente, was andersartige digitale Objekte jedoch nicht ausschließt. Wissenschaftliche Repositories stellen das Rückgrat des so genannten Grünen Weges zu Open Access dar (siehe [1]).

Gegenstandsbereich eines *Institutional Repository* (IR) sind die elektronischen Publikationen oder andere digitale Objekte, die durch die Mitarbeit von Mitgliedern einer – zumeist wissenschaftlichen – Institution entstanden sind. Das Institutional Repository einer solchen Einrichtung gehört demnach zu deren Informationsinfrastruktur. Es bildet eine Art digitales Schaufenster, das den wissenschaftlichen Output der Institution – zumindest teilweise – nach außen darstellt. Mit dieser Prämisse sollten die Überlegungen zum Aufbau eines eigenen IR beginnen, der im Folgenden schrittweise beschrieben wird.

*Der Artikel beschreibt in Form eines instruktiven Leitfadens mögliche Herangehensweisen beim Aufbau eines Institutional Repository und stellt Vorüberlegungen und Entscheidungsoptionen dar, die sich für Vorbereitung und Betrieb einer Publikationsplattform ergeben. Die Definition der Anforderungen und die darauf aufbauende Argumentation erfolgen auf der Basis des DINI-Zertifikats für Dokumenten- und Publikationsservices.*

*Institutional Repository, DINI-Zertifikat, IR-Software, Leitlinien, Nachhaltigkeit.*

<sup>1</sup> siehe <http://roar.eprints.org/> [09.02.2009]

<sup>2</sup> siehe <http://www.dini.de/wiss-publizieren/repository/> [09.02.2009]

## 1. Schritt: Klärung grundlegender Fragestellungen

Zunächst ist es wichtig, Klarheit über folgende Fragestellungen in Bezug auf das aufzubauende Institutional Repository zu gewinnen:

1. Welchem übergreifenden Zweck soll das IR dienen?
2. Wer soll welche Inhalte über das IR publizieren?
3. An welche potentielle Leserschaft soll sich das IR richten?
4. Wie soll sich das IR in andere Aktivitäten im Bereich der Informationsbereitstellung integrieren?

Anhand der Antworten sollte auch versucht werden, das jährlich zu erwartende Publikationsvolumen abzuschätzen und notwendige Maßnahmen zur Unterstützung von Autoren sowie weitere bereitzustellende Dienstleistungen und technische Schnittstellen abzuleiten.

Der Klärung der Frage, welche Publikationen in Form von Open Access veröffentlicht werden sollen, sollte besondere Aufmerksamkeit geschenkt werden. Ihr liegen prinzipielle Erwägungen zugrunde – etwa eine mögliche Open-Access-Strategie der jeweiligen Einrichtung. Sie muss aber im Einzelfall vielfach in Kooperation mit den jeweiligen Autoren und vor allem mit den potentiellen Herausgebern abgewogen werden.

Zu den ersten Aufgaben beim „Neustart“ eines IR gehört es außerdem, dieses organisatorisch sinnvoll zu verankern und mit entsprechenden technischen und personellen Kapazitäten auszustatten. Auch hierfür ergeben sich wichtige Hinweise aus der Beantwortung der oben genannten Fragen.

### Zum Beispiel: Die HU

Sollen als Autoren – wie beispielsweise an der Humboldt-Universität – grundsätzlich alle Wissenschaftlerinnen und Wissenschaftler bzw. Studierende, die gleichzeitig Angehörige der jeweiligen Hochschule sind, angesprochen werden, ist ein großes Spektrum an unterschiedlichen Publikationsarten und eine Vielzahl von Wissenschaftsdisziplinen zu berücksichtigen. Die Betreuung dieser durch sehr hetero-

gene Anforderungen geprägten Veröffentlichungen muss sich entsprechend auch in den Leitlinien für das Institutional Repository wiederfinden. Einen dergestalt konzipierten Dokumentenservers innerhalb der zentralen Einrichtungen der Hochschule anzusiedeln, liegt nahe und verlangt zentral von der Universität gesteuerte Ressourcen.

Sobald diese grundlegenden Fragen geklärt sind und die organisatorische Verortung des IR innerhalb der Einrichtung vollzogen ist, kann mit der praktischen Realisierung begonnen werden. Besonders hilfreich hierfür sind die Hinweise und Empfehlungen der Deutschen Initiative für Netzwerkinformation (DINI, siehe Kasten).

## Das DINI-Zertifikat

Die 1998 gegründete Deutsche Initiative für Netzwerkinformation, ein Zusammenschluss von Rechenzentren, wissenschaftlichen Bibliotheken, Medienzentren und Vertretern aus dem Wissenschaftsbereich, vergibt seit 2004 das so genannte DINI-Zertifikat für Dokumenten- und Publikationsservices. In seiner aktuellen Fassung aus dem Jahr 2007 beschreibt das DINI-Zertifikat Qualitätsstandards für die zu verwendenden Technologien, die organisatorische Einbettung und den Betrieb von IRs einschließlich der dazugehörigen Dienstleistungen [2]. Die Lektüre der Anforderungen, die im DINI-Zertifikat als Mindeststandards beschrieben werden, bietet sich als hilfreiche Orientierung für den Aufbau eines IR an. Sie kann darüber hinaus als Entscheidungsgrundlage dafür dienen, ob der Service tatsächlich an der eigenen Institution eingerichtet werden oder im Rahmen einer externen Kooperation durch einen Dritten erbracht werden soll. Gerade für kleinere Institutionen ist das technologische Hosting eines IR häufig überlegenswert (vgl. [3]).

Innerhalb Deutschlands bilden die im DINI-Zertifikat festgelegten Anforderungen unter anderem die Grundlage für die Vernetzung der einzelnen IRs sowie den Aufbau darauf basierender übergreifender Dienste. Beispielsweise arbeitet das durch die Deutsche Forschungsgemeinschaft geförderte Projekt *Open-Access-Netzwerk* auf eine solche Vernetzung deutscher IRs hin und zielt dabei auf eine verbesserte nationale und internationale Wahrnehmung der lokal angebotenen Publikationen und damit des Forschungsbeitrags insgesamt (siehe [4]).

## 2. Schritt: Integration des IR in die strategische Planung

Im nächsten Schritt ist es wichtig, für das Vorhaben die Unterstützung innerhalb der oberen Entscheidungsebenen der Institution zu gewinnen. Das ist unter anderem erforderlich, um die Nachhaltigkeit des aufzubauenden Repository zu gewährleisten und dessen Betrieb für die nächsten Jahre und Jahrzehnte finanziell abzusichern. Dabei ist zu berücksichtigen, dass vor allem Personalkapazitäten langfristigen Planungen und Verträgen unterliegen. Aber auch die regelmäßige Erneuerung der Hard- und Software für das IR verursacht Folgekosten. Durch die strategische Planung wird die Einbettung des IR in die Informationsinfrastruktur der Einrichtung und darüber hinaus gewährleistet. Es muss an fachliche und biblio-

thekarische Netzwerke sowie an weitere Aktivitäten zur Unterstützung der digitalen Forschung und Lehre angebunden werden.

## 3. Schritt: Formulierung von Leitlinien für das IR

Die Leitlinien für ein Institutional Repository dienen der Dokumentation der Zielstellungen, Grundsätze und Rahmenbedingungen. Sie stellen damit öffentlich dar, wie die im ersten Schritt dieses Artikels formulierten Grundsatzfragen für das jeweilige IR beantwortet werden. Im DINI-Zertifikat ist eines der Kriterien dieser Problematik gewidmet und es wird detailliert beschrieben, was in einer solchen *Policy* enthalten sein sollte. Die Leitlinien stellen sowohl für potentielle

Autoren als auch für die Nutzer der sonstigen Dienstleistungen des IR klar, auf welcher Basis es betrieben wird und welche Erwartungen sie in Bezug auf Verfügbarkeit, Zuverlässigkeit und Qualität der Dienste haben dürfen.

Gemäß DINI-Zertifikat muss die Policy für ein IR folgende Punkte beinhalten:

1. eine Beschreibung der Zielgruppe und Aussagen darüber, wer auf dem Repository publizieren darf,
2. eine Definition von Verfahren zum Betrieb des IR und die Festlegung von Leistungen, die der Anbieter des Publikationsservice Autoren und Herausgebern bereitstellt,
3. die Festlegung der inhaltlichen, funktionalen und technischen Qualität der auf dem IR veröffentlichten Dokumente im Sinne einer Beschränkung (beispielsweise auf wissenschaftliche Abhandlungen oder Veröffentlichungen, die einen Peer-Review-Prozess durchlaufen haben),
4. eine Definition von Archivierungszeiträumen und die Garantie, die Dokumente in Abhängigkeit ihrer inhaltlichen, funktionalen und technischen Qualität für einen definierten Zeitraum nutzbar anzubieten,
5. eine Beschreibung der Rechte und Pflichten von Autoren und Herausgebern,
6. eine Erklärung zu Open Access, die das jeweilige Verständnis des Begriffs in der Institution ebenso enthält wie einrichtungsspezifische Handlungsrichtlinien für Autoren (ggf. bezogen auf bestimmte Publikationsarten) und vorhandene Umsetzungsinstrumente.

#### 4. Schritt: Bereitstellung von Personal

Entscheidend für die erfolgreiche Realisierung eines Institutional Repository an einer wissenschaftlichen Einrichtung ist nicht allein die verwendete Technologie, sondern auch das zur Verfügung stehende Personal. Das Ziel muss darin bestehen, ein Team zusammenzustellen, das alle notwendigen Fähigkeiten und Fertigkeiten zum Aufbau, zur Wartung und zum Betrieb des IR und der damit ver-

bundenen Dienstleistungen abdeckt. Zumindest bibliothekarische und Informatik-Kompetenzen sollten im Team vertreten sein. Darüber hinaus sollte der Bereich der Öffentlichkeitsarbeit bzw. des Marketing fachlich abgedeckt sein.

Abgesehen davon ist der Aspekt der Nachhaltigkeit zu berücksichtigen. Befristete und fremdfinanzierte Projektmitarbeiter sind für die Konzeption und den unmittelbaren Aufbau eines IR und den damit verbundenen Service durchaus hilfreich und oftmals alternativlos. Allerdings sollte bereits zu Beginn eines solchen Vorhabens bedacht werden, wie Betrieb und Wartung des IR nach Ablauf einer Startphase mit unbefristet tätigem Personal abgedeckt werden können und das entsprechend notwendige Know-how weitergegeben werden kann, um die Stabilität des entstandenen Dienstes auch langfristig sicherzustellen.

#### 5. Schritt: Auswahl der Technologie

Für die Auswahl der geeigneten Technologie ist entscheidend, die gängigen Anforderungen, die an ein IR und die darauf basierenden Funktionen und Dienstleistungen gestellt werden, zu berücksichtigen und in bestmöglicher Weise zu erfüllen. Auf der Basis internationaler Entwicklungen und anerkannter Standards hat DINI für die technologische Basis von Repositories Mindeststandards erarbeitet, die im bereits erwähnten DINI-Zertifikat festgeschrieben sind (siehe Kasten auf Seite 50).

Zu den allgemeinen Aspekten, die zur Bewertung von IR-Software herangezogen werden (siehe z. B. [5], [6]), gehören unter anderem die Einfachheit der Installation und der Wartung des Systems, die Möglichkeit, mehrere Installationen gleichzeitig innerhalb einer Hard- und Softwareumgebung laufen zu lassen (relevant beispielsweise für Hosting-Services, siehe [3]), die Interoperabilität mit anderen Systemen und das Vorhandensein entsprechender Schnittstellen, um das IR in die Informationsinfrastruktur der Einrichtung und darüber hinaus zu integrieren. Beispielsweise mag es aus Sicht einer konkreten Einrichtung von Bedeutung

sein, das IR in den OPAC zu integrieren, in Lehr- und Lernumgebungen einzubinden oder bei der Authentifizierung auf das lokale Nutzermanagement (LDAP) zurückzugreifen. Weitere Aspekte sind die Skalierbarkeit und die Sicherheit des Systems, die Größe der Entwickler- und Nutzergemeinde der Software, die jeweiligen Lizenzbedingungen für die Nachnutzung sowie die Möglichkeiten, eigene Veränderungen und Erweiterungen des Quellcodes vorzunehmen.

Die meisten der heute verwendeten Lösungen sind ursprünglich aufgrund einer am konkreten Bedarf einer einzelnen Institution ausgerichteten Entwicklung entstanden. In einigen Fällen ist daraus später ein nachnutzbares und an lokale Erfordernisse anpassbares Produkt geworden, dessen Weiterentwicklung inzwischen zumeist in den Händen einer größeren Entwicklergemeinschaft liegt und das die üblichen Zyklen einer Softwareentwicklung durchläuft. Inzwischen existiert eine ganze Reihe technischer Lösungen für Institutional Repositories, die sofort einsetzbar sind und die Mehrzahl der technischen Anforderungen des DINI-Zertifikats bereits erfüllen. Im nichtkommerziellen Bereich sind dabei international vor allem EPrints, DSpace und Fedora sowie innerhalb Deutschlands zusätzlich OPUS und MyCoRe zu nennen.

Spätestens hier stellt sich die Frage, welches der zur Verfügung stehenden Produkte ausgewählt werden sollte. Dies hängt nicht nur von den bereits genannten allgemeinen bzw. durch das DINI-Zertifikat gegebenen Anforderungen ab, sondern muss mit Berücksichtigung der lokalen Gegebenheiten und Erfordernisse entschieden werden, beispielsweise in Abhängigkeit der Art des jeweiligen Betreibers und der bereitzustellenden Publikationen. Beispielsweise werden Repositories nicht nur von Universitätsbibliotheken oder Rechenzentren betrieben, sondern als fachspezifische Publikationsplattformen zuweilen auch durch Fachbereiche, Fachgesellschaften oder außeruniversitäre Forschungseinrichtungen. Daher sind auch die Ansprüche, die hinsichtlich organisatorischer und technischer Fragestellungen an die Software gestellt werden, sehr unterschied-

## Anforderungen an ein IR gemäß DINI-Zertifikat

(Die numerischen Referenzen beziehen sich auf die Strukturierung des Textes im DINI-Zertifikat [2].)

- **WWW:** Es muss eine Webschnittstelle für Nutzer vorhanden sein (Punkt 2.6.3).
- **Upload:** Die Bereitstellung von Online-Formularen für Autoren zum Upload ihrer Publikationen (Document upload tools), ist erforderlich (Punkt 2.3).
- **Rechte:** Der Autor kann eigene Publikationen über entsprechende Vereinbarungen mit Lizenzen versehen, z. B. CC, DPPL (Punkt 2.4).
- **Nachweis:** Die Dokumente müssen aus technischer Sicht kontrolliert und nachweisbar aufgenommen werden (Punkt 2.5.1).
- **Persistent Identifiers:** Die Verwendung von Persistent Identifiers ist erforderlich, z. B. URN, DOI (Punkt 2.5.2).
- **Keywords:** Jede Publikation sollte mit aussagekräftigen Schlagwörtern zumindest in deutscher und englischer Sprache versehen sein. Auf technologischer Ebene muss daher eine verbale Sacherschließung durch freie Schlagwörter oder ein Klassifikationssystem unterstützt werden (Punkt 2.6.1).
- **DDC:** In Ergänzung dazu ist die Anwendung der Dewey-Dezimalklassifikation (DDC) gemäß der verwendeten Sachgruppen in der Deutschen Nationalbibliografie erforderlich (Punkt 2.6.1).
- **Metadaten:** Die Metadaten müssen frei zugänglich angeboten werden und zumindest gemäß Dublin Core (ISO 15836:2003) strukturiert sein (Punkt 2.6.2).
- **OAI:** Das IR muss über eine OAI-Schnittstelle verfügen, die entsprechend den OAI-Empfehlungen von DINI realisiert ist (Punkt 2.6.2).
- **Statistik:** Jedes IR muss im Rahmen der rechtlichen Bestimmungen eine eigene konsistente Zugriffsstatistik führen (Webserver-Log) (Punkt 2.7).
- **Langzeitarchivierung:** Die Dokumente und ihre Metadaten müssen dauerhaft miteinander verbunden sein. Für die Langzeitarchivierung der Publikationen sollte die Kooperation mit einer anerkannten Archivierungsinstitution, wie z. B. der Deutschen Nationalbibliothek (DNB), gesucht werden (Punkt 2.8).
- **Dokumentation:** Es muss eine Dokumentation des technischen Systems mit relevanten Versionsangaben und technischen Parametern zu allen Komponenten existieren (Punkt 2.5.1).

lich. Die Wahl einer geeigneten Software hängt insbesondere auch davon ab, welche Anforderungen aus Sicht der jeweiligen Betreibereinrichtung tatsächlich Priorität besitzen. Zudem haben sich speziell in Deutschland Lösungen etabliert, die auf die spezifischen Belange hiesiger Universitätsbibliotheken zugeschnitten sind und somit im internationalen Vergleich wenig Berücksichtigung finden.

Der Rahmen des vorliegenden Artikels erlaubt zwar eine detaillierte Betrachtung der einzelnen Systeme und eine Analyse der jeweiligen Vor- und Nachteile nicht (siehe dazu beispielsweise [7] und [8]). Einige Anhaltspunkte zu den genannten Produkten für Institutional Repositories finden sich jedoch in den folgenden Abschnitten sowie im nebenstehenden Kasten (Seite 51).

### OPUS

Die ursprünglich als *Online-Publikationssystem der Universität Stuttgart* entwickelte Software existiert seit 1998 und diente zunächst ausschließlich der Bereitstellung elektronischer Hochschulschriften an dieser Universität. Mit seinen bundesweit mehr als 60 Installationen ist OPUS heute die am weitesten verbreitete IR-Software in Deutschland und Grundlage für Hosting-Angebote mehrerer Bibliotheksverbünde. Durch die Ausrichtung auf deutsche Hochschulen unterstützt OPUS die wichtigsten hier erforderlichen Schnittstellen und Standards – etwa die Generierung von URNs, deren maschinenlesbare Bereitstellung im Format EPICUR zur Abfrage durch die Deutsche Nationalbibliothek (DNB) und eine den

Anforderungen des DINI-Zertifikats entsprechende OAI-Schnittstelle, die neben Dublin Core auch das für die Meldung an die DNB erforderliche Metadatenformat XMetaDiss beherrscht. Darüber hinaus stellt OPUS ein Modul zur Definition eigener Nutzungslizenzen (u. a. Creative Commons) bereit, aus denen beim Upload von Publikationen ausgewählt werden kann.

OPUS unterstützt vor allem den Publikationsprozess in der Bibliothek: Nach der Einstellung eines Dokuments durch dessen Autor wird es durch einen Mitarbeiter bezüglich technischer und formaler Kriterien – etwa Vollständigkeit der Publikation und Korrektheit der Metadaten – überprüft, ggf. korrigiert und anschließend freigeschaltet. Die Publikationen können durch Browsing nach unterschiedlichen Kategorien sowie mit einer Suche in den Metadaten und den Volltexten (über Google) aufgefunden werden. Abgesehen von der Open-Access-Einstellung erlaubt es die Definition differenzierter Zugriffsbeschränkungen für Dokumente und ermöglicht darüber hinaus die Verwaltung von Schriftenreihen und Sammlungen mit beliebig gestuften Hierarchien.

OPUS ist ein robustes und vielfach erprobtes System, dessen Entwicklergemeinschaft jedoch sehr abgeschlossen ist und dessen Anpassungsmöglichkeiten – beispielsweise in Bezug auf das Datenmodell und die grafische Oberfläche – sehr beschränkt sind.

### DSpace

*DSpace*, am Massachusetts Institute of Technology (MIT) in Kooperation mit der Firma HP entwickelt, befindet sich beim MIT seit April 2004 im Produktionsbetrieb. Die stetige Weiterentwicklung erfolgt inzwischen über die *DSpace Foundation*, in der sich weitere große Forschungseinrichtungen zusammengeschlossen haben, sowie parallel durch Einzelinstitutionen, da der Quellcode – als Open Source zur Verfügung steht. So hat beispielsweise die Universitätsbibliothek Kassel das System so erweitert, dass es auch das URN-Format der DNB unterstützt. In der Standardversion unterstützt DSpace bereits die gängigen Schnittstellen und Standards wie OAI,

## IR-Software im Überblick

### OPUS

**Entwickler:** Universität Stuttgart  
**Verbreitung:** ca. 60 Installationen in Deutschland  
**Lizenz:** Universität Stuttgart, 250 € einmalige Gebühr  
**Programmiersprache:** PHP  
**Systemumgebung:** Linux/UNIX, Apache, MySQL  
**Download:** Per Mail über opus@uni-stuttgart.de  
**Support:** Wiki, Bug-Tracker

### DSpace

**Entwickler:** DSpace Foundation  
**Verbreitung:** ca. 550 Installationen weltweit, davon 7 in Deutschland  
**Lizenz:** Open Source (BSD)  
**Programmiersprache:** Java, JSP  
**Systemumgebung:** Linux/UNIX bzw. Windows, Apache, Tomcat, PostgreSQL bzw. Oracle  
**Download:** [http://sourceforge.net/project/showfiles.php?group\\_id=19984](http://sourceforge.net/project/showfiles.php?group_id=19984)  
**Support:** DSpace Federation, Bug-Tracker, Wiki, FAQ, User Group Meetings

### EPrints

**Entwickler:** University of Southampton  
**Verbreitung:** ca. 280 Installationen weltweit, davon ca. 10 in Deutschland  
**Lizenz:** Open Source (GPL)  
**Programmiersprache:** Perl  
**Systemumgebung:** Linux/UNIX bzw. Windows, Apache, MySQL  
**Download:** <http://www.eprints.org/software/>  
**Support:** Wiki, Dateiarchiv, Mailingliste

### MyCoRe

**Entwickler:** Hochschulübergreifendes Entwicklerteam  
**Verbreitung:** ca. 20 Installationen in Deutschland  
**Lizenz:** Open Source (GPL)  
**Programmiersprache:** Java  
**Systemumgebung:** Linux/Unix bzw. Windows oder Mac OS, Jetty bzw. Tomcat oder Websphere, relationale Datenbank (hsq, MySQL, Oracle, DB2)  
**Download:** [http://sourceforge.net/project/showfiles.php?group\\_id=92005](http://sourceforge.net/project/showfiles.php?group_id=92005)  
**Support:** Mailingliste, Bug-Tracker, Wiki, IRC-Kanal

### Fedora

**Entwickler:** FedoraCommons  
**Verbreitung:** 150 Installationen weltweit, davon ca. 5 in Deutschland  
**Lizenz:** Open Source (Mozilla Public Licence)  
**Programmiersprache:** Java  
**Systemumgebung:** Linux/Unix, Apache,  
**Download:** [http://sourceforge.net/project/showfiles.php?group\\_id=177054](http://sourceforge.net/project/showfiles.php?group_id=177054)  
**Support:** Bug-Tracker, Wiki, Mailinglisten, Forum

Handle und OpenURL. Außerdem können Creative-Commons-Lizenzen vergeben werden, und es erlaubt nutzerseitig das Browsing über einzelne Sammlungen sowie eine Suche in den Metadaten.

Bei Installation, Anpassung und Wartung des Systems werden Java-Kenntnisse benötigt. Die Einrichtung gestaltet sich etwas komplexer als beispielsweise im Falle von OPUS. Aus technologischer Sicht ist das System wegen der zugrunde liegenden Java-Technologien dafür zukunftssträchtiger, da bei der Programmierung und Erweiterung strukturierter vorgegangen werden kann. DSpace stellt ein ausgereiftes Nutzermanagement bereit, bei dem der Publikationsworkflow rollenbasiert gesteuert wird. Das System arbeitet bezüglich Autorisierung, Urheber- und Verwertungsrechten und Zugriff mit nutzergruppenspezifischen Anpassungen. Es ist skalierbar, gut für größere Forschungseinrichtungen nutzbar und natürlich einsetzbar, um den Grünen Weg zu Open Access zu unterstützen.

### EPrints

*EPrints* ist eine Software der University of Southampton aus dem Jahr 2000 und wird durch das dortige Entwicklerteam auch weiterhin gewartet und herausgegeben. Weltweit existieren fast 300 produktive Installationen des Systems. Im Gegensatz zu den anderen Produkten ist EPrints für die durch Wissenschaftler vollzogenen Eigenpublikationen – das so genannte Self-Archiving – optimiert, wird jedoch auch zum Betrieb von IRs im hier verwendeten Sinne eingesetzt. EPrints unterstützt den Publikationsprozess für Autoren, Herausgeber und Institutionen. Es verfügt über ein dafür ausgelegtes Rechtemanagement. Das System stellt eine OAI-Schnittstelle bereit und erlaubt standardmäßig den Export von Metadaten, unter anderem in den Formaten Dublin Core, MODS, METS, BiBTeX und OpenURL. Eine URN-Unterstützung und eine Alerting-Funktion, die über die Bereitstellung neuer Dokumente im Repository informiert, sind ebenfalls implementiert. Die History-Funktion ermöglicht Administratoren das Verfolgen von Änderungen an Dokumenten und Metadaten. Um EPrints installieren zu können,



werden keine Programmierkenntnisse benötigt. Sie werden erst für Erweiterungen oder tief greifende Anpassungen erforderlich. Grundsätzlich sind mehrere Archive unter einer Installation lauffähig, das heißt, ein Hosting ist prinzipiell möglich. Die Standarddistribution ist für den englischsprachigen Betrieb optimiert und muss daher bei der Installation für deutschsprachige Nutzer entsprechend angepasst werden. Anpassungen sind auch bei der OAI-Schnittstelle notwendig, um den Export von Metadaten in den Formaten EPICUR und XMetaDiss zu ermöglichen.

## MyCoRe

MyCoRe, kurz für *My Content Repository*, wurde im Rahmen des Projekts MILESS an der Universität Essen (heute Universität Duisburg-Essen) entwickelt und steht inzwischen als Open-Source-Software zur Verfügung, die durch ein hochschulübergreifendes Konsortium erweitert und gepflegt wird. MyCoRe bildet dabei einen Rahmen, der mehrere Softwaremodule und ein übergreifendes Management-System integriert, mit dem digitale Sammlungen unterschiedlichster Art verwaltet werden können. Das System stellt alle wesentlichen Grundfunktionen eines digitalen Repository und zur Unterstützung des Publikationsprozesses bereit – angefangen von einer Uploadschnittstelle, über die Sacherschließung einschließlich Klassifikation bis hin zur individuellen Anpassung der Oberfläche. Anpassungen für einfache Anwendungen können auf der Konfigurationsebene vorgenommen werden und bedingen keinen Eingriff in den Programmcode. Besonders hervorzuheben ist die Möglichkeit, das Datenmodell vollständig konfigurieren zu können, hierbei kommen XML-basierte Konfigurationsdateien zum Einsatz. Dadurch ist das System nicht auf die Verwaltung bestimmter Publikationsarten beschränkt. Außerdem stellt MyCoRe wichtige Schnittstellen und Standards wie URN, OAI, XMetaDiss, epicur, Z39.50, MAB2 und SOAP bereit. Es unterstützt Video-Streaming ebenso wie eine Suche in den Metadaten und den Volltexten – unter anderem auch für verteilte MyCoRe-Installationen. Das System ist für den deutschen Sprachraum konzipiert und

kann damit ohne sprachliche Anpassungen unmittelbar eingesetzt werden. Für den produktiven Einsatz von MyCoRe mit einer langfristigen Perspektive sollten personelle Ressourcen mit Java-Kompetenzen vorhanden sein.

## Fedora

Die ursprünglich an der University of Virginia und der Cornell University entwickelte Repository-Software *Fedora* (kurz für *Flexible Extensible Digital Object and Repository Architecture*) stellt im Gegensatz zu den anderen vorgestellten Systemen kein unmittelbar einsetzbares und in sich abgeschlossenes Komplettsystem bereit. Fedora bildet eher eine Basisarchitektur, die über ein flexibles Framework Dienste bereitstellt, auf deren Grundlage ein Repository implementiert werden kann. Das Softwarekonzept folgt dem Ansatz der Serviceorientierten Architektur (SOA), die sich durch gute Modularisierbarkeit und hohe Flexibilität auszeichnet. Das System ist in Java implementiert, die einzelnen digitalen Objekte werden intern auf der Basis von RDF (Resource Description Framework) abgebildet und beschrieben. Für Import und Export von Objekten stehen das Metadatenframework METS sowie das selbst entwickelte Format *Fedora Object XML* (FOXML) zur Verfügung. Im Sinne der bereits erwähnten SOA verfügt Fedora lediglich über einige zentrale Webservices, die über das Protokoll SOAP angesprochen und genutzt werden können.

Mit diesen Schnittstellen stellt die Kernsoftware damit nur die Basisdienste zur Verfügung, die notwendig sind, um auf die digitalen Objekte zuzugreifen und wenige Funktionen daran auszuführen. Alle anderen Komponenten, insbesondere auch die eigentliche Benutzerschnittstelle, müssen auf dieser Basis selbst implementiert werden. Andererseits lässt sich Fedora dafür wie keines der anderen Systeme in bestehende Umgebungen und Anwendungen integrieren.

Im Rahmen des Projekts *eSciDoc* arbeitet die *Max Planck Digital Library* (MPDL) daran, ein vollständiges IR-System auf der Basis von Fedora zu entwickeln und es anschließend als nachnutzbares Produkt zur Verfügung zu stellen.

## Wer die Wahl hat ...

Alle hier vorgestellten Systeme genügen – ggf. nach entsprechender Konfiguration – den im DINI-Zertifikat formulierten technologischen Anforderungen. Sie kommen also potentiell alle als Grundlage für ein Institutional Repository in Betracht. Entscheidend für die Auswahl eines der Systeme sollten neben den funktionalen Unterschieden der einzelnen Lösungen vor allem die lokalen Gegebenheiten und Anforderungen sein. Dies betrifft nicht nur die Flexibilität und Anpassbarkeit der einzelnen Produkte – beispielsweise in Bezug auf das zugrunde liegende Datenmodell und das Layout der Benutzeroberflächen – sondern insbesondere auch die an der betreffenden Einrichtung vorhandenen Kompetenzen und Systemumgebungen.

## 6. Schritt: Etablieren von Workflows

Auf den ersten Blick scheint es vielleicht so, dass Institutional Repositories, einmal vorbereitet und eingerichtet, sozusagen ohne weiteren Aufwand arbeiten bzw. einfach durch Autoren und Herausgeber gefüllt werden, die Veröffentlichungen in der entsprechenden inhaltlichen und technischen Qualität über die Uploadschnittstelle einreichen. Doch in der Regel sind Institutional Repositories keine reinen Selbstbedienungssysteme. Die einzelnen Veröffentlichungsvorgänge werden zwar sozusagen von außen angestoßen, müssen aber in die bibliothekarischen Geschäftsgänge integriert werden. Genau hierin besteht eine der zentralen Herausforderungen beim Aufbau eines IR. Dazu sollten bestehende Arbeitsabläufe vergegenwärtigt und explizit beschrieben, neue Anforderungen spezifiziert und daraus resultierende Geschäftsprozesse definiert und dokumentiert werden.

Dabei ist die Online-Bereitstellung von Zweitveröffentlichungen im Sinne des Grünen Weges zu Open Access noch der einfachste Fall: Hier erfolgt seitens des Betreibers für hochgeladene Publikationen allenfalls eine Plausibilitätsprüfung und ggf. die Korrektur von Metadaten, bevor sie freigeschaltet und somit

öffentlich verfügbar gemacht werden. Mehrteilige Publikationsvorhaben, bei denen als Gegenüber des IR ein Herausgeber in Erscheinung tritt, bedingen oftmals weit komplexere Geschäftsgänge – beispielsweise durch die Unterstützung des gesamten Dokumenterstellungsprozesses einschließlich der Erzeugung eines formatierten Satzspiegels bzw. eines XML-basierten Dateiformats als Grundlage für einen medienneutralen Publikationsansatz. Und bei der Veröffentlichung von Hochschulschriften spielen die gültigen Prüfungsordnungen mit ihren Publikationsbestimmungen in die Workflows mit hinein, sodass möglicherweise die Ausstellung entsprechender Bescheinigungen und die Herstellung gedruckter Exemplare mit in den Aufgabenbereich des Publikationsservice fallen.

Um eine möglichst hohe Qualität der bibliografischen Daten zu erreichen, ist die Erfassung der Metadaten oder zumindest deren Nachbearbeitung durch bibliothekarisches Personal empfehlenswert. Dies gilt insbesondere dann, wenn die Daten in andere Systeme exportiert werden sollen – beispielsweise in den lokalen OPAC oder in einen Verbundkatalog. Für die Sacherschließung müssen – so sieht es auch das DINI-Zertifikat vor – entsprechende Richtlinien vorhanden sein, in denen beschrieben ist, welche Erschließungsinstrumente verwendet werden und welche Regelungen dabei gelten.

Grundsätzlich sollten Autorenverträge geschlossen werden – je nach erforderter Rechtsverbindlichkeit in schriftlicher Form, wie etwa bei Dissertationen, oder zumindest durch entsprechend zu bestätigende Klauseln im Online-Formular, beispielsweise bei Pre- und Postprints.

Überaus sinnvoll erscheint die Integration von Geschäftsprozessen rund um das IR in die Erstellung einer Institutsbibliografie, wie sie an vielen wissenschaftlichen Einrichtungen gepflegt wird. IR und Jahresbibliografie haben grundsätzlich denselben Gegenstandsbereich – Publikationen von Angehörigen der jeweiligen Institutionen (siehe z. B. [9]). Durch die Integration beider Systeme können nicht nur Synergieeffekte genutzt werden, sie kann auch die Motivation von Wissenschaftlern erhöhen, ihre Ver-

öffentlichungen in das IR einzustellen – beispielsweise dann, wenn die Anzahl eigener Veröffentlichungen für die leistungsbezogene Mittelvergabe herangezogen und diese Kennzahl anhand der Institutsbibliografie ermittelt wird.

Für komplexe Geschäftsgänge, die unterschiedliche Benutzerrollen und Aktivitäten beinhalten – wie beispielsweise inhaltliche Begutachtungsprozesse – ist die Verwendung von Workflowsystemen, die die Prozesse elektronisch unterstützen, empfehlenswert.

## 7. Schritt: Kundenbetreuung

Neben dem technischen Betrieb des Institutional Repository muss auch die Betreuung der Autoren und Herausgeber – also der Kunden des IR – bewerkstelligt werden. Dazu zählt beispielsweise, Fragen zu Dokumentformaten zu beantworten und bei rechtlichen Fragen – vor allem im Zusammenhang mit dem Urheberrecht und bestehenden Autorenverträgen mit Dritten – auf entsprechende Informationsangebote zu verweisen bzw. die Informationen selbst zu beschaffen – beispielsweise mithilfe der SHERPA/RoMEO-Liste<sup>3</sup>, in der Open-Access-Regelungen zahlreicher Wissenschaftsverlage hinterlegt sind (siehe dazu [10]).

Das DINI-Zertifikat erwartet als Mindestanforderung im Bereich der Autorenbetreuung öffentlich verfügbare Kontaktinformationen und ein Beratungsangebot – beispielsweise per Online-Formular, FAQ-System, E-Mail- oder Telefon-Hotline. Es liegt auf der Hand, dass der Betreuungsbedarf mit komplexer werdenden Dienstleistungen gegenüber Herausgebern und Autoren steigt. Werden beispielsweise Dokumentvorlagen für Texterstellungssysteme zur Verfügung gestellt, um bei komplexeren Publikationsvorhaben die Erstellung eines einheitlichen Satzspiegels bzw. eines XML-basierten Dokumentworkflows vorzubereiten, sind auch dafür entsprechende Beschreibungen bzw. Unterstützungsangebote vorzuhalten, z. B. in Form regelmäßig durchgeführter Kurse.

Besondere Beachtung gebührt Her-

ausgebern von Schriftenreihen, Sammelbänden, Zeitschriften und dergleichen, die ihre Publikationsvorhaben über ein IR veröffentlichen wollen. Hierbei gestaltet sich die Betreuung als kooperativer Beratungsprozess, der unter anderem die Berücksichtigung individueller Anforderungen – beispielsweise an Funktionalität oder Layout – zum Inhalt hat (siehe dazu auch [11]).

## 8. Schritt: Öffentlichkeitsarbeit und Vernetzung

Wenn die Grundsatzfragen geklärt, Leitlinien erarbeitet, Ressourcen und Personal organisiert sind, ein System installiert ist, Workflows definiert und Betreuungsangebote geschaffen sind, fehlen nur noch – die Autoren. Dass institutionelle Veröffentlichungsplattformen keine Selbstläufer sind, zeigen die relativ geringen Publikationszahlen in nahezu allen vorhandenen IRs. Gründe dafür sind der zusätzliche – wenn auch geringe – Aufwand, den Autoren für die Einstellung von Zweitpublikationen in Repositories aufbringen müssen und der demgegenüber für sie nicht erkennbare Mehrwert sowie andererseits die geringe Akzeptanz gegenüber Open-Access-Publikationen unter Wissenschaftlern (siehe [12]).

Neben der Bereitstellung brauchbarer Mehrwertdienste für die potentielle Leserschaft einerseits und der Unterstützung von Autoren und der Minimierung des erforderlichen Aufwands andererseits (siehe 7. Schritt) lässt sich diesem weit verbreiteten Problem nur mit der Schaffung externer (finanzieller) Anreize und wirkungsvollem Marketing begegnen. Während finanzielle Anreize für die Nutzung von IRs nicht ohne verbindliche und durch die jeweiligen Führungsebenen zu erlassenden Regelungen möglich sind – vor allem Rahmen leistungsbezogener Mittelvergabe – liegt die Öffentlichkeitsarbeit eher in der Hand der Betreiber von IRs. Dabei sollte keine Möglichkeit ausgelassen werden, potentielle Autoren nicht nur auf das bestehende Publikationsangebot, sondern auch auf die Vorzüge dieser Veröffentlichungsform hinzuweisen – angefangen von einer möglichst prominenten Verlinkung des IR von den

<sup>3</sup> siehe <http://www.sherpa.ac.uk/romeo/> [09.02.2009]

zentralen Internetseiten der Einrichtung, über Aushänge, Informationen per E-Mail oder Hauspost, Artikel und Anzeigen in internen Publikationen bis hin zu Informationsveranstaltungen. Als sinnvoll erweist es sich dabei, auf bestehende Kommunikations- und Organisationsstrukturen zurückzugreifen und potentielle Autoren und Herausgeber darüber anzusprechen – etwa über die Fachreferenten der Bibliotheken oder durch die gezielte Ansprache von Wissenschaftlern durch den Besuch von deren Arbeitsgruppenbesprechungen und Kolloquien. Wirkungsvolle Öffentlichkeitsarbeit ist ein personalintensives Arbeitsfeld und idealerweise eng mit der – sich im Erfolgsfall anschließenden – Kundenbetreuung verzahnt.

Letztlich geht es darum, mit dem IR für die Institution und deren Mitglieder identitätsstiftend zu wirken und zu erreichen, dass Wissenschaftler und Studierende das Repository als ihre Publikationsplattform begreifen – und zwar zum einen in der Rolle potentieller Autoren, aber auch als lesende Nutzer. Insofern ist die Erhöhung der Bekanntheit nicht nur zur Gewinnung von Autoren, sondern auch für die allgemeine Akzeptanz des IR als zentraler Bestandteil institutioneller Informationsversorgung zu verstehen.

Wichtig sind parallel dazu die Vernetzung des IR innerhalb der nationalen und internationalen Community und die Einspeisung der Metadaten in übergreifende – möglicherweise fachbezogene – Recherchedatenbanken. Wenn das IR die durch das DINI-Zertifikat beschriebenen Anforderungen erfüllt, sind dafür die technischen Voraussetzungen im Sinne der Interoperabilität mit entsprechenden Serviceanbietern bereits gegeben. Damit die bibliografischen Daten tatsächlich Eingang in die einschlägigen Nachweissysteme finden und die auf dem IR vorgehaltenen Publikationen über das eigene Webangebot hinaus sichtbar und wahrgenommen werden, ist die Registrierung des IR und der OAI-Schnittstelle bei zentralen Diensten und in der offiziellen Registratur der *Open Archives Initiative* erforderlich<sup>4</sup>. Die meisten OAI-basierten Service Provider übernehmen regelmäßig

die dort vorhandenen Informationen über vorhandene IRs, um dann von ihnen die verfügbaren Metadaten einzusammeln. Darüber hinaus ist aber auch die Anmeldung des IR bei weiteren Verzeichnissen wie dem bereits genannten ROAR und OpenDOAR<sup>5</sup> sinnvoll.

Nicht zuletzt sollte nach Abschluss aller Arbeiten das DINI-Zertifikat auch formal beantragt werden. Einmal erlangt, kann es als positiver Marketingeffekt auf potentielle Autoren und Herausgeber wirken. Es ist beispielsweise auch die Voraussetzung, um mit voller Funktionalität in das nationale Portal von OA-Netzwerk eingebunden zu werden (siehe [4]) und damit auch die Tür zu dessen internationalem Pendant – zu DRIVER<sup>6</sup>.

## Fazit

Institutional Repositories gehören zur Informationsversorgung in wissenschaftlichen Einrichtungen längst zum Alltag. Damit sie durch deren wesentliche Zielgruppe – die jeweiligen Wissenschaftler in der Rolle als Autor, Herausgeber und Leser – auch tatsächlich akzeptiert und selbstverständlich genutzt werden, müssen noch viele Hürden genommen werden. Eine wesentliche Voraussetzung ist ein überlegtes und strukturiertes Vorgehen bei Konzeption und Aufbau von IRs und der Planung der Dienstleistungen und Prozesse. Dabei muss das sprichwörtliche Rad heute nirgendwo mehr neu erfunden werden – weder in Bezug auf die benötigte Technologie noch hinsichtlich der strategischen Überlegungen und der notwendigen organisatorischen Maßnahmen. Dennoch sollten die Besonderheiten der jeweiligen Einrichtungen und die daraus resultierenden spezifischen Anforderungen bei der konkreten Umsetzung nicht unberücksichtigt bleiben. Die Orientierung an den Bedürfnissen der potentiellen Nutzer kann ein wesentlicher Erfolgsfaktor für das gesamte IR sein.

## Literatur

- [1] SCHIRMBACHER, P.; MÜLLER, U.: *Der „Grüne Weg zu Open Access“ – Aktuelle Entwicklungen in Deutschland*. cms-journal 32, 2009, S. 41–46.
- [2] DINI E.V.: *DINI-Zertifikat für Dokumenten- und Publikationsservices 2007*, DINI-Schriften 3, 2007, urn:nbn:de:kobv:11-10079197 [09.02.2009].
- [3] DOBRATZ, S.: *Hostingservice für andere Dokumentenserver*. cms-journal 32, 2009, S. 86–89.
- [4] MALITZ, R.: *Open Access – Verfügbar ist nicht präsent*. cms-journal 32, 2009, S. 70–73.
- [5] OPEN SOCIETY INSTITUTE: *OSI Guide to Institutional Repository Software, Version 3*, 2004, [http://www.soros.org/openaccess/pdf/OSI\\_Guide\\_to\\_IR\\_Software\\_v3.pdf](http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf) [Stand 02.07.2007].
- [6] WYLES, R.: *Technical Evaluation of Research Repositories*. Open Access Repositories in New Zealand. 2006, <https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf> [09.02.2009].
- [7] DOBRATZ, S.: *Open-Source-Software zur Realisierung von Institutionellen Repositorien – Überblick*. Zeitschrift für Bibliothekswesen und Bibliografie 54 (4-5). 2007, S. 199–206.
- [8] ROGGENBUCK, R.; SCHMITZ, J.: *Vortrag auf dem DINI-WS am 19.02.2009 in Göttingen*.
- [9] ABEL, G.; DRESSLER, B.; HERB, U.; MÜLLER, M.: *Verzahnung von Jahresbibliographie und Institutional Repository der Universität des Saarlandes*. Bibliotheksdienst 40 (6), 2006, S. 726–733.
- [10] ZIELKE, D.: *Rechtliche Probleme bei Postpublikationen und die Ergebnisse des DFG-Projekts „Open Access Policies“*. cms-journal 32, 2009, S. 60–62.
- [11] MÜLLER, U.: *Auf dem goldenen Weg zum edoc-Server – Open-Access-Zeitschriften, ...* cms-journal 32, 2009, S. 29–32.
- [12] OVER, A.; MAIWORM, F.; SCHELEWSKY, A.: *Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access*. Studie der Deutschen Forschungsgemeinschaft. Weinheim: Wiley-VCH, 2005.

<sup>5</sup> siehe <http://www.opendoar.org/> [09.02.2009]

<sup>6</sup> siehe <http://www.driver-repository.eu/> [09.02.2009]

<sup>4</sup> siehe <http://www.openarchives.org/Register/BrowseSites> [09.02.2009].

# Sieben auf einen Streich – Argumente für eine Open-Access-Publikation auf dem edoc-Server der HU

Es gibt gute Gründe, warum Sie Ihre wissenschaftlichen Beiträge als Erst- oder Zweitveröffentlichung auf dem Dokumenten- und Publikationsserver (edoc-Server) der Humboldt-Universität zu Berlin platzieren sollten:

## 1. Sichtbarkeit: Hier werden Sie gelesen!

Im World-Wide-Web wird besser gefunden, wer professionell publiziert. Der edoc-Server ermöglicht durch die Indexierung mit Metadaten das Auffinden Ihrer Publikation durch Suchmaschinen, wie sie etwa von Fachportalen oder Virtuellen Bibliotheken eingesetzt werden. [Lesen Sie dazu den Artikel Malitz: Open Access – Verfügbar ist nicht präsent, S. 70]

## 2. Standardisierung: Zertifizierter Server mit internationalen Standard

Der edoc-Server ist als Institutional Repository in weltweite Netzwerke eingebunden und unterstützt internationale Standards, die beständig weiterentwickelt werden. Der Dokumenten- und Publikationsserver der Humboldt-Universität besitzt seit 2004 das DINI-Zertifikat. [Lesen Sie dazu den Artikel Dobratz/Müller: Wie entsteht ein Institutional Repository? S. 47]

## 3. Kostenpunkt: Kein finanzieller Aufwand

Der Open-Access-Gedanke zielt vorrangig auf den barriere- und kostenfreien Zugang zu wissenschaftlicher Information. Zugleich wird damit Autoren ermöglicht, ohne finanziellen Aufwand zu publizieren. Der edoc-Server erhebt keinerlei Gebühren. [Lesen Sie dazu den Artikel Gradmann: Publizierens im Open Access Modell, S. 20]

## 4. Urheberrecht: Weltweit akzeptierte Lizenzmodelle

Der edoc-Server garantiert Ihnen die konsequente Wahrung Ihrer Rechte an den Publikationen. Als Autor können Sie selbst bestimmen, inwieweit Sie Ihre Publikation anderen zur Verfügung stellen und welche Rechte Sie ihnen einräumen. [Lesen Sie dazu die Artikel Klatt-Kafemann: Creative Commons-Lizenzen, S. 81; Zielke: Open Access Policies, S. 60]

## 5. Langzeitarchivierung: Nicht im Netz verschwinden

Das Internet vergisst schnell. Digitale Datenbestände müssen gepflegt und an technologische Innovationen angepasst werden, der edoc-Server bietet Ihnen die Sicherheit, Ihre Publikation dauerhaft auffindbar und lesbar zu erhalten. [Lesen Sie dazu den Artikel Dobratz: Grundfragen der digitalen Langzeitarchivierung für den edoc-Server, S. 93]

## 6. Service: Von Metadaten bis Print on Demand

Als Dokumenten- und Publikationsserver bietet Ihnen der edoc-Server zahlreiche Dienste, die Nutzung und Wirkung ihrer Publikation zu verbessern, etwa durch ein Datenformat zum Import der Metadaten in Literaturverwaltungssysteme oder durch einen Print-on-Demand-Dienst. Zudem können Sie anhand der Zugriffszahlen die Nutzung Ihrer Publikationen nachvollziehen. [Lesen Sie dazu den Artikel: Service für Autorinnen und Autoren, S. 58]

## 7. Zeitmanagement: Schneller Zugang und kompetente Autorenbetreuung

Ihre wissenschaftliche Arbeit wird ohne Anlauf- oder Wartezeiten publiziert. Als Autor können Sie im Falle einer Erstveröffentlichung die bereitgestellten Dokumentvorlagen benutzen oder im Falle der Zweitveröffentlichung eines Aufsatzes bzw. Buchkapitels Ihren Beitrag in digitaler Form über eine Schnittstelle hochladen. Beim gesamten Publikationsprozess werden Sie vom Team des edoc-Servers unterstützt. [Lesen Sie dazu den Artikel: Autorenhinweise, S. 102]

# eSciDoc – Das Repository-Konzept der Max Planck Digital Library

Malte Dreyer | malte.dreyer@mpdl.de

Ulla Tschida | tschida@mpg.de

*Basierend auf der eSciDoc Infrastruktur entwickelt die MPDL ein Set an Services und Anwendungen, die als Repository für unterschiedliche Forschungsdaten, u.a. Publikationsdaten, in der MPG eingesetzt werden. Der Beitrag skizziert ein möglichst umfassendes Bild der Ausgangslage, der fachlichen und technischen Anforderungen sowie des gewählten Entwicklungsansatzes. Zusätzlich wurde ein Schwerpunkt auf die organisatorischen und sozialen Aspekte eines Infrastruktur-Projektes für eScience in einer großen Wissenschaftsorganisation gelegt.*

*eSciDoc, MPDL, FIZ, Repository, Research data, eScience, Infrastruktur, eResearch Environment*

## Einleitung

eSciDoc ist ein vom BMBF gefördertes gemeinsames Projekt der Max-Planck-Gesellschaft (MPG)<sup>1</sup> und dem Fachinformationszentrum FIZ Karlsruhe<sup>2</sup>. Innerhalb der MPG ist das Projekt an der zum 1. Januar 2007 gegründeten Max Planck Digital Library (MPDL)<sup>3</sup> angesiedelt. Das Ziel des Projekts ist die Entwicklung einer disziplinübergreifenden virtuellen Forschungsumgebung im Rahmen der eScience-Initiative des Bundes. Die Infrastruktur und darauf aufbauende Anwendungen stehen unter [www.escidoc.org](http://www.escidoc.org) im Rahmen einer Open-Source-Lizenz zur Verfügung. Innerhalb der MPG werden die entwickelten Anwendungen zurzeit für den Produktiveinsatz vorbereitet, außerhalb der MPG evaluieren international über 20 Einrichtungen einen möglichen Einsatz oder entwickeln bereits erste eigene Anwendungen und Services. Die weitere Entwicklung und Pflege von Infrastruktur und Anwendungen wird auch über das Ende der Förderung zum 31. Juli 2009 hinaus durch eigene Kapazitäten von MPG und FIZ Karlsruhe sichergestellt.

Kernanwendungen, die auf der Infrastruktur aufsetzen, sind zurzeit ein System zum Publikationsdatenmanagement „PubMan“, ein System zur kollaborativen Arbeit mit Bild-Daten „FACES“ sowie ein System zur Verwaltung und Anreicherung von digitalisierten Text-Ressourcen „ViRR“.

1 Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., <http://www.mpg.de>

2 Fachinformationszentrum FIZ Karlsruhe, <http://www.fiz-karlsruhe.de>

3 Max Planck Digital Library, <http://mpdl.mpg.de>

## Ausgangslage – eScience-Herausforderungen

### Heterogenität der Ressourcen

Durch die zunehmende Datenzentriertheit der Forschungsaktivitäten steigt der Bedarf an technischen Infrastrukturen, die unterschiedliche digitale Ressourcen und Artefakte der wissenschaftlichen Arbeit transparent bearbeitbar und nachhaltig verfügbar halten. Für diesen Zweck werden zurzeit meist einzelne dedizierte Anwendungen mit speziell entwickelter Datenhaltung entwickelt. Zunehmend virtuelle, globale und interdisziplinäre Forschungsansätze erfordern jedoch die Identifikation und Zugänglichkeit von Ressourcen unterschiedlicher Herkunft sowie Möglichkeiten zur Anreicherung, zur Verbindung und Integration von Ressourcen und die Anwendung von disziplinspezifischen Werkzeugen, um die Ressourcen in neuen Kontexten zu nutzen.

### Technologiegestützte Kommunikation

Kommunikation von Ergebnissen ist eine grundlegende Bedingung für wissenschaftlichen Fortschritt. Wurden bislang die Ergebnisse selektiv und aggregiert in Form einer wissenschaftlichen Publikation veröffentlicht, ermöglichen zunehmend webbasierte Forschungsaktivitäten eine transparente und durchgehende Dokumentation der einzelnen Arbeitsschritte, z. B. auf welcher Grundlage gearbeitet wurde (i. e. Qualität der verfügbaren digitalen Ressourcen bzw. Entitäten), wie die Ausgangslage bearbeitet oder modifiziert wurde (i. e. Qualität der Vorgehensweisen



und Methoden) und welche (Zwischen-) Ergebnisse für die weitere Forschung verwandt wurden<sup>4</sup>. Die bewusste Zugänglichkeit der wissenschaftlichen Artefakte und ihrer jeweiligen Kontexte ermöglicht somit neuartige und vor allem unerwartete Ansetzungspunkte für neue Forschungsfragen. Ressourcen sind somit ebenso einem stetigen Änderungsprozess unterworfen wie auch die entsprechenden begleitenden Informationen zur Herkunft, zur inhaltlichen Beschreibung oder zu den Nutzungskontexten. eScience-Anwendungen unterstützen diese kommunikativen Prozesse und müssen deshalb als Arbeitsumgebungen wissenschaftszentriert konzipiert und gestaltet werden. Der Begriff „Usability“ erhält somit eine umfassende Bedeutung als wissenschaftsnahe Gestaltung von Arbeitsabläufen zur Suche und Identifikation relevanter Wissensbestände im Netz, zur transparenten Beschreibung von angewandten Ressourcen, Methoden und Ergebnissen sowie zum Einsatz kooperativer Werkzeuge zum Austausch von Artefakte und über Artefakte.

## Semantic Web

Begleitend zu diesen Entwicklungen stehen die informationswissenschaftlichen Konzepte des „Semantic Web“, um Ressourcen für diese Vorhaben günstig aufzubereiten und angereichert bereitzustellen. „The coolest thing to do with your data will be thought of by someone else“<sup>5</sup> umschreibt das Potential sowie die Herausforderung an maschinenlesbare Semantik von Daten, ihren Kontexten und Inhalten. Technologien wie Ontologien, Collection-, Service- oder Metadaten-Registrierungen können hier die Identifikation und Nutzbarkeit von relevanten Entitäten wie Daten, Dokumenten, Personen, Organisationen oder Konzepten erhöhen.

<sup>4</sup> Nicht nur innerhalb der wissenschaftlichen Community, sondern verstärkt im öffentlichen und politischen Kontext, sind die Forderungen nach Kommunikation von wissenschaftlichen Ergebnissen spürbar: Förderorganisationen, Industrie, Nicht-Regierungsorganisationen, Politik, die Zivilgesellschaft und andere soziale Gruppen erwarten eine transparente Kommunikation von wissenschaftlichen Tätigkeit und Ergebnissen.

<sup>5</sup> Leitspruch der JISC Common Repository Interfaces Group CRIG <http://www.ukoln.ac.uk/repositories/digirep/index/CRIG>

## Technische Anforderungen

Um diese nur grob skizzierte Ausgangslage zu adressieren, wurden in der Entwicklung der eSciDoc-Infrastruktur verschiedene technische Anforderungen a priori definiert, die im Laufe der Entwicklung einzelner Anwendungen konsolidiert werden.

### Entitäten, Versionierung

Die Verwaltung von digitalen Ressourcen beinhaltet nicht nur die reine Ablage von Bits und Bytes, sondern auch die entsprechenden Zusatzinformationen bzw. Metadaten in ihrem Bezug zu den Ressourcen, die zu einem beliebigen Zeitpunkt nachvollziehbar abgebildet werden müssen. Diese komplexeren Anforderungen werden durch entsprechende Versionierungssysteme, wie z. B. „Subversion“, nicht direkt unterstützt und erfordern eine breitere Sicht auf mögliche Informationseinheiten, ihre jeweiligen Komponenten und Metadaten.

### Zugriffsrechte

Darüber hinaus sind im Kontext von Forschungsaktivitäten Ressourcen teilweise frei zugänglich, teilweise mit sehr eng gefassten Zugriffsrechten verbunden, wobei sich die Zugriffsrechte auch während eines Lebenszyklus verändern können. Um vielfältige Szenarien möglichst feingranular, auch in dezentralen Authentifizierungsmodellen, abbilden zu können, müssen entsprechende Technologien unterstützt werden.

### Interoperabilität

Zum Austausch von Daten mit anderen Systemen, zur Anbindung bestehender Anwendungen oder disziplinspezifischer Werkzeuge ist ein breites Spektrum an disziplin- oder technologiespezifischen Schnittstellen sowie die Verwendung von standardisierten Protokollen und offenen Formaten erforderlich. Ebenso sind vielfältige Sichten und Präsentationsformen von Ressourcen und deren Zusammenstellungen zu unterstützen und in der Infrastruktur formal als Content Models zu beschreiben.

## Content Models

Die Quantität und Heterogenität wissenschaftlich relevanter Ressourcen erfordert die Möglichkeit, auch grundsätzlich unbekannte Inhalte und entsprechende Metadaten – quasi in einer Black Box – abbilden zu können. Um isolierte Datensilos zu vermeiden, muss die Infrastruktur möglichst umfassende Informationen zum Kontext des Inhalts sowie zur Art der Teilkomponenten verstehen. Dafür ist eine formale semantische Beschreibung von Inhaltstypen in Form von Content Models erforderlich.

### Objektmodell

Ein erstes abstraktes Objektmodell muss die Anforderungen an Versionierung, persistenter Identifikation, Relationen und Annotationen sowie Autorisierung erfüllen. Im Zuge der Implementierung verschiedener Anwendungen wird das Objektmodell verfeinert, um die disziplin-, institutions- und ressourcenspezifischen Charakteristika zu unterstützen.

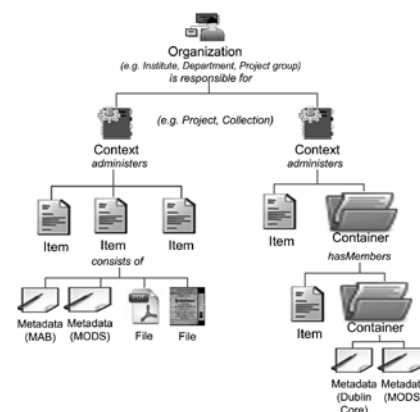


Abb.1 eSciDoc Objektmodell

## Digital Preservation

Die längerfristige Speicherung binärer Datenströme sowie die technischen Anforderungen an Sicherheit, Emulation und Migration der Datenbestände sind von den zuständigen Einrichtungen bereits adressiert. Neben der Erhaltung der reinen Daten sind Informationen zur Identifikation und Semantik von Beständen, deren Kontext und Lebenszyklus für eine spätere Verwendung erforderlich.

Unabhängig von den organisatorischen Aspekten der Ablage von Forschungsdaten unterstützt die eSciDoc Infrastruktur die technische Langlebigkeit der Datenhaltung, da für sämtliche Daten die entsprechenden Informationen zu Kontext, Eigentümern, Historie bzw. Lebenszyklus in offenen und standardisierten Formaten abgelegt werden.

## Vorgehen im eSciDoc Projekt

Basierend auf der skizzierten Ausgangslage, den daraus resultierenden technischen Anforderungen sowie den besonderen Zielsetzungen innerhalb der MPG ergaben sich für das eSciDoc-Projekt, neben den technischen Aspekten, bestimmte Prämissen, die die Projektkultur, Arbeitsweisen und Projektergebnisse beeinflussten.

### Forschungs- und Ressourcenzentriert

Sämtliche Entwicklungen gehen auf konkrete, wissenschaftlich basierte Szenarien einzelner Fachdisziplinen zurück. Die Entwicklung der eSciDoc-Anwendungen erfordert eine zeitnahe und intensive Auseinandersetzung mit den disziplinspezifischen Artefakten, Arbeits- und Kommunikationsprozessen, weshalb Anwendungen im engen und ständigen Austausch mit den Nutzenden konzipiert werden.

Zu den grundlegenden Fragestellungen in der Phase der Konzeption und des Designs einer Anwendung zählen Typologie und Charakteristika der Artefakte sowie notwendiger und sinnvoller Granularität der Informationseinheiten bzw. ihrer Aggregationen. Im Zusammenhang des Lebenszyklus von bestimmten Ressourcen spielt die Definition von offenen und geschützten Artefakten eine wichtige Rolle. Das Verständnis für gewohnte Kommunikationskanäle und -partner ermöglicht die Definition von Anforderungen an Auffindbarkeit und Interoperabilität, im Besonderen im Hinblick auf potentielle interdisziplinäre Nachnutzung in bekannten und neuen Kontexten (re-use und re-purposing). Ein Kernstück des wissenschaftlichen Arbeitsablaufes sind angewandte Metho-

den und Werkzeuge, um Artefakte zu generieren, zu erschließen, zu bearbeiten und zu publizieren. Dabei stellt sich oft die Herausforderung, bereits bestehende disziplinspezifische, oft proprietäre Werkzeuge in offene Systeme zu integrieren.

### Interaktion von Anwendung und Infrastruktur

Die Entwicklung von forschungszentrierten Anwendungen steht in ständigem Austausch mit der parallelen Entwicklung der Infrastruktur. Beide Ebenen bedingen einander und dienen dem gegenseitigen Ausbau: Auf der fachlichen Ebene werden Arbeitsprozesse, Artefakte und Werkzeuge identifiziert sowie auf mögliche generische und disziplinspezifische Komponenten hin untersucht. Komplementär werden auf der technischen Ebene generische Services und disziplinspezifische Solutions entwickelt. Wie bei jedem Infrastrukturprojekt ist die Herausforderung gegeben, an „Rohbau“ und „Innenausbau“ gleichzeitig zu arbeiten und die resultierenden Implikationen rechtzeitig und zeitnah zu erkennen.

Derzeit werden drei eSciDoc-basierte Anwendungen entwickelt, die jeweils verschiedene Szenarien und Artefakte unterstützen. Die Anwendung PubMan adressiert das Management und die Dissemination von Publikationsdaten. Komplementär dazu werden zwei Lösungen für das Szenario einer „Scholarly Workbench“ entwickelt: FACES beinhaltet den nachhaltigen Umgang und die Arbeit mit Bilddaten, VIRR adressiert die nachhaltige Ansichts- und Editions Umgebung für digitalisierte Textressourcen.<sup>6</sup> Alle drei Anwendungen haben eine solution-spezifische Logik, greifen jedoch auf generische Services zurück (z. B. import service).

<sup>6</sup> Weitere Angaben zu den einzelnen Anwendungen finden sich im CoLab Wiki der MPDL [http://colab.mpdl.mpg.de/mediawiki/ESciDoc\\_Solutions\\_summary](http://colab.mpdl.mpg.de/mediawiki/ESciDoc_Solutions_summary)

## Open Source & Community Building

Die Entwicklungen des eSciDoc-Projekts stehen sämtlich als Open Source frei zur Verfügung. Als Lizenz wurde die CDDL gewählt. Nach vier Jahren intensiver Community-Arbeit innerhalb der MPG, der Identifikation und Konsolidierung von Early Adoptern, Piloten, Partnern und Mentoren auf allen Ebenen der MPG, wendet sich das eSciDoc-Projekt nun dezidiert der internationalen Community zu, um die Ergebnisse für neue Einsatzmöglichkeiten und Szenarien zur Verfügung zu stellen sowie die Nachhaltigkeit der Entwicklung zu gewährleisten.<sup>8</sup> Das Projekt muss sich seitdem stärker den Aspekten stabiler Interfaces, einem transparenten Entwicklungsvorgehen, den Fragestellungen zur Einbindung externer Entwickler und der stetigen Verbesserung der allgemeinen Kommunikation stellen. Die zahlreichen Anfragen von nationalen und internationalen Organisationen an Einsatz- und Entwicklungsmöglichkeiten von eSciDoc<sup>9</sup> erfordern ein entsprechendes Management der Erwartungshaltungen bzw. umfassender und vorausschauender Entwicklungsplanung. Als Ziel für die eSciDoc Days in 2009 steht deshalb eine verbesserte Transparenz in der Planung und Organisation von speziellen Anwendungsgruppen. Die diversen Anwendungsinteressen der MPG und der externen Interessenten werden in gesonderten Arbeitsgruppen oder „Special Interest Groups“ (z. B. zur Bilddissemination) adressiert und gemeinsam bearbeitet.

<sup>7</sup> <http://www.sun.com/cddl/>. Als Lizenz mit beschränktem Copyleft sichert sie zum einen die weitere freie Verfügbarkeit ab, schränkt jedoch nicht die Erstellung von eingeschränkt nutzbaren Lösungen ein, die z. B. Patentrechte berühren. Ein weiterer Vorteil der CDDL ist die Möglichkeit zur freien Benennung eines Gerichtsstandorts, ohne hierbei eine neue Lizenz erzeugen zu müssen, wie dies z. B. bei der MPL (<http://www.mozilla.org/MPL/MPL-1.1.html>) der Fall ist.

<sup>8</sup> Als Startschuss zum Aufbau einer Open Source Community wurden im Juni 2008 die eSciDoc Days mit über 100 internationalen Teilnehmenden in Berlin abgehalten. Hierbei konnten durch die unterschiedlichen Ideen und Vorstellungen der Teilnehmenden zu einem Einsatz von eSciDoc neue Szenarien identifiziert werden, für welche die Infrastruktur eingesetzt werden kann.

<sup>9</sup> Zur Zeit evaluieren in etwa 20 internationale Organisationen den Einsatz von eSciDoc.

## Zusammenfassung und Rückblick

Der gewählte Ansatz für den Aufbau einer strategischen Infrastruktur hat sich bislang hinsichtlich gewählter Technologien, Transparenz in der Planung und enger Zusammenarbeit mit den Instituten als geeignet für die Anforderungen der MPG gezeigt. Das strategische Ziel, eine nachhaltige eScience-Umgebung aufzubauen, steht als Top-Down Entscheidung in einem potentiellen Konflikt mit den gewohnt Bottom-Up Ansätzen in wissenschaftlicher IT-Entwicklung. Umso dringender ist es, bewusste Balance zwischen schnellen, sichtbaren Mehrwertdiensten für einzelne Fachdisziplinen und den langfristigen, oft nicht deutlich wahrnehmbaren Investitionen in Nachhaltigkeit und Effizienz einer Infrastruktur zu halten. Die oft unterschiedlichen Anforderungen der Organisation und der Anwender erfordern ein sorgfältiges und diplomatisches Management der Erwartungen sowie transparente Projektplanung. Der gewählte Ansatz, frühzeitig interessierte und vor allem engagierte „Early Adopters“ zu finden, die als Fürsprecher der einzelnen Disziplinen auftreten, war dabei sehr hilfreich. Die frühe und fokussierte Einbindung der Fachdisziplinen sowie anderer Stakeholder ermöglicht parallel den Aufbau einer „Knowledge Infrastruktur“, die eine Begleiterscheinung einer eScience-Organisation ist: Kompetenzen und Expertise über Artefakte, wie z. B. Bilddaten, sind über die Disziplinen als auch über traditionelle Expertiseträger (wie Wissenschaftler, IT und Bibliotheken) verteilt und es erfordert entsprechende Netzwerkarbeit, um implizites und explizites Wissen zusammenzubringen.

Die Anforderungen des Projekts erfordern entsprechend flexible interne Prozesse zum Projektmanagement und Expertise-Aufbau: vom Team ist nicht nur ein „Abarbeiten“ von Anforderungen gefordert, sondern vor allem ein „Erarbeiten“ von Expertise in fachlichen und technischen eScience-Fragestellungen zur Ableitung einzelner Dienste. Die unterschiedlichen Ziele und Vorgehensweisen für notwendige analytische und evaluative Konzeptions- und Architekturfragen und „harte“ Terminierung fer-

tiger, reifer Produkte erfordern ein hohes Maß an intrinsischer Motivation und konstruktiver Mitarbeit des Teams. Die Projektplanung selbst muss von beiden Partnern flexibel genug gestaltet werden, um das Potential von unerwarteter, aber sinnvoller Nutzung der Infrastruktur und ihrer Komponenten auszuschöpfen: „be prepared for the unexpected“.

Nur kurz soll an dieser Stelle auch auf die Probleme der Komplexität größerer Software-Entwicklungsvorhaben hingewiesen werden. Im Projekt waren zeitweise über 20 Personen mit der reinen Softwareentwicklung beschäftigt, was eine feingranulare Planung der einzelnen Schritte, eine systemunterstützte Prozesssteuerung sowie strikte Vereinbarungen zu den Softwareentwicklungsumgebungen erfordert. Gleichzeitig hat der Aufbau dieser formalen Strukturen einen robusten softwaretechnischen Grundstein für die nun immer verteilte Entwicklung gelegt.

Unabhängig von der strategischen Zielsetzung und den konkreten Anforderungen einer Organisation stellt sich für Entscheidungsträger die Frage der Nachhaltigkeit größerer Infrastrukturprojekte. Ein Projekt wie eSciDoc bietet durch den erheblichen Ressourceneinsatz in der Aufbauphase einen guten Ausgangspunkt für weiteren Ausbau und Entwicklungen, erfordert jedoch auch Beiträge vieler unterschiedlicher Institutionen im Sinne eines Open-Source-Community-Ansatzes, um die Einsatzmöglichkeiten und die Servicevielfalt zu steigern.

Die vielleicht deutlichste Herausforderung von Infrastrukturprojekten ist der implizite Bedarf an paralleler Organisationsentwicklung, die eigenen, oft nicht direkt beeinflussbaren Mechanismen unterliegt. IT-Systeme modellieren und greifen somit in bestehende Abläufe ein, das gilt umso mehr für übergreifende Infrastrukturprojekte, die sämtliche Bereiche einer Organisation berühren. Sowohl technische als auch organisatorische und soziale Strukturen müssen neu gedacht und konzipiert werden, was sich am deutlichsten bei Themen wie Digital Curation, persistenter Identifikation, Authentication/Authorisation oder Implementierung von Open-Access-Workflows zeigt. Hier stellt sich nicht

nur die Frage nach Technologien, sondern auch die Frage nach nachhaltigen und kosteneffizienten Lösungen für das Zusammenspiel von zentralen und lokalen bzw. fachspezifischen Systemen, Expertisen und nicht zuletzt Personal- und Sachmitteln.

# Rechtliche Probleme bei Postpublikationen und die Ergebnisse des DFG-Projekts „Open Access Policies“

Dennis Zielke | zielkede@cms.hu-berlin.de

Der kostenfreie Zugang zu wissenschaftlichen Informationen im Internet, „Open Access“, wird u. a. durch die zusätzliche Bereitstellung von bereits in einem Verlag publizierten oder noch zu publizierenden wissenschaftlichen Arbeiten, das sogenannte Self-Archiving von Pre- bzw. Postprints, erreicht.

Diese Form der Veröffentlichung führt zu einer erhöhten Sichtbarkeit der Open-Access-Dokumente, die zur Reputation der einzelnen Autoren bzw. ihrer Institution beitragen. Open-Access-Beiträge sind zudem für jeden Interessenten barrierefrei zugänglich. [1] Essenziell ist in diesem Zusammenhang die umfassende Information des Autors darüber, welche Position der Verlag zur Strategie des Self-Archivings bezieht.

Das Vertragsverhältnis zur Publikation wird grundsätzlich zwischen Autor und Verlag geschlossen. Im Vertragswerk sind urheberrechtliche Rahmenregelungen verankert, wobei beide Vertragsnehmer innerhalb dieses Rahmens freien Gestaltungsspielraum haben. Dabei sind die Autoren gegenüber den Verlagen aber zumeist in der schwächeren Position, da sie weder die Zeit noch die Kenntnisse besitzen, sich mit den detaillierten Einzelheiten des Verlagsvertrages auseinanderzusetzen.

Hierbei unterstützt die Arbeitsgruppe Elektronisches Publizieren [2] von Computer- und Medienservice und der Universitätsbibliothek der Humboldt-Universität zu Berlin die Autoren bereits seit über 10 Jahren in Fragen zur Klärung von Veröffentlichungsrechten und bei der Aufbereitung der zu publizierenden Dokumente. Durch die Möglichkeit einer Veröffentlichung auf dem edoc-Server, dem Institutional Repository der Humboldt-Universität zu Berlin, kann die Publikation unter dem Autorennamen, dem einzelnen Fachgebiet sowie dem Institut gefunden werden. Außerdem ist die Publikation über den edoc-Server an internationale Fachdatenbanken angeschlossen und dadurch für ein Fachpublikum recherchierbar.

Um die Verhandlungsposition der einzelnen Autoren bei der Ausgestaltung des Vertrages zu stärken, wurde im Jahr 2002 an der Universität Nottingham das Projekt RoMEO ins Leben gerufen, mit dessen Hilfe entsprechende Informationen von Verlagen im englischsprachigen Raum gesammelt und dann klassifiziert werden. Im Projekt SHERPA [3] wird die Datenbasis seit dem Jahr 2003 weiter gepflegt. Darin wurden Zeitschriftentitel der British Library aufgenommen und mit bereits bestehenden Verlagsinformationen verknüpft.

*„Was gestatten deutsche Wissenschaftsverlage ihren Autoren“ – ist eine häufig gestellte Frage bei Postpublikationen auf Publikations- und Dokumentenservern. Mit dem DFG-Projekt „OA-Policies“ werden dazu Antworten gegeben.*

*Open Access, Autoren, Open Access Policies, Publikation*

## Publisher copyright policies & self-archiving

Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

### Search

Search ☒ Journal titles ☐ Publisher names

for

find ☐ contains or ☒ starts with or ☐ Exact phrase only ☐ ISSN

### Browse

- [All publishers](#)
- [Green publishers](#)
- [Blue publishers](#)
- [Yellow publishers](#)
- [White publishers](#)

[What the colours mean.](#)

### RoMEO News

- [Latest Updates to RoMEO](#)
- [Publishers Allowing use of their PDFs in Repositories](#)
  - [Press Release](#) - 27-Aug-2008
  - [List of Publishers](#)
- [Publishers with Paid Options for Open Access](#)
  - [Press Release](#) - 03-Sept-2008
  - [List of Publishers](#)

### Recently Added Publishers



- [Mohr Siebeck](#) - 23-Dec-2008
- [Digital Information Research Foundation](#) - 23-Dec-2008
- [Old City Publishing](#) - 23-Dec-2008
- [Psychopharmacology & Biological Narcology](#) - 22-Dec-2008
- [Society of American Foresters](#) - 17-Dec-2008

Abb. 1: Eingabemaske für die SHERPA/RoMEO-Datenbank (engl\_client.png)

## Das Projekt „Open Access Policies“ – Technische Aspekte

Um die englischsprachige Datenbank mit starker Dominanz englischer Verlage auch für deutsche Verlage, Autoren und Bibliotheken besser nutzbar zu machen, wurde das von der Deutschen Forschungsgemeinschaft geförderte Projekt „Open Access Policies“ ins Leben gerufen. Dieses Projekt beschäftigt sich mit den Rechten, die die deutschen Verlage ihren Autoren gestatten und schafft Rahmenbedingungen im Umgang mit Open Access.

Dabei wurde im Verlaufe der Projektphase die RoMEO-Datenbank mit den Richtlinien deutscher Verlage und um eine deutschsprachige Schnittstelle [4] erweitert.

Somit können die wissenschaftlichen Autoren beim Einbringen ihrer Dokumente herausfinden, ob und zu welchen Bedingungen ein Verlag eine frei zugängliche Kopie einer Publikation erlaubt.

Die SHERPA/RoMEO-Datenbank hält ein Application Programming Interface (API) bereit, welches über einen GET-Request aufgerufen wird und als Resultat einen XML-String zurückliefert.

Momentan werden die XML-Daten vom deutschen Client zur Laufzeit übernommen und inhaltlich unverändert ausgegeben. Damit diese Daten in weitere Sprachen übersetzt und inhaltlich ergänzt werden können, hat das „Open Access Policies“-Projekt die erweiterte Schnittstelle „oaPAPI“ entwickelt, die es gestattet, beliebig viele Datenquellen und unterschiedliche Datenstrukturen einzulesen und in einem gemeinsamen XML-String zu kombinieren.

Um einen autorisierten Nutzerkreis über einen leicht bedienbaren Zugang an der Datenpflege beteiligen zu können, nutzt das Projekt zur Speicherung eigener Daten das Kollaborationswerkzeug „Mediawiki“. Damit ist es möglich geworden, die Informationen der RoMEO-Liste u. a. um Quellenangaben und theoretisch unbegrenzt viele Sprachalternativen zu erweitern.

In der weiteren Entwicklung wird das „Open Access Policies“-Projekt ein Werkzeug anbieten, das die Daten des SHERPA/RoMEO-APIs um weitere durch das Projekt gesammelte Informationen und Sprachvarianten erweitert und seinerseits ein mit den zusätzlich hinzukommenden Daten „angereichertes“ XML-API ausgibt. Dadurch wird es möglich, die Datensätze in mehreren Sprachen anzuzeigen. In der derzeitigen Projektphase wird daran gearbeitet, die bereits entwickelte Schnittstelle zu Ro-

...opening access to research

## SHERPA RoMEO deutsch

### Was gestatten Verlage bei der Selbstarchivierung im Open Access?

[Startseite](#) | [Anleitung](#) | [Eingabeformular](#) | [Kontakt](#) | [SHERPA-RoMEO](#)

Diese Datenbank soll helfen, die Standardbedingungen der Verlage bei Open Access-Publikationen für Autoren transparenter zu machen.

#### Recherche

Suche nach ☐ Zeitschriften ☒ Verlagen

Suchworte

Ergebnisse finden über ☐ irgendein Suchwort, ☒ alle Suchworte oder ☐ den exakten Ausdruck

#### Vorauswahl

Alle Verlage

☒ Grüne Verlage ☐ Blaue Verlage ☐ Gelbe Verlage ☐ Weiße Verlage

[Was bedeuten die Farben?](#)

Abb. 2: Eingabemaske für den deutschen SHERPA/RoMEO Client (sherpa\_romeo.png)



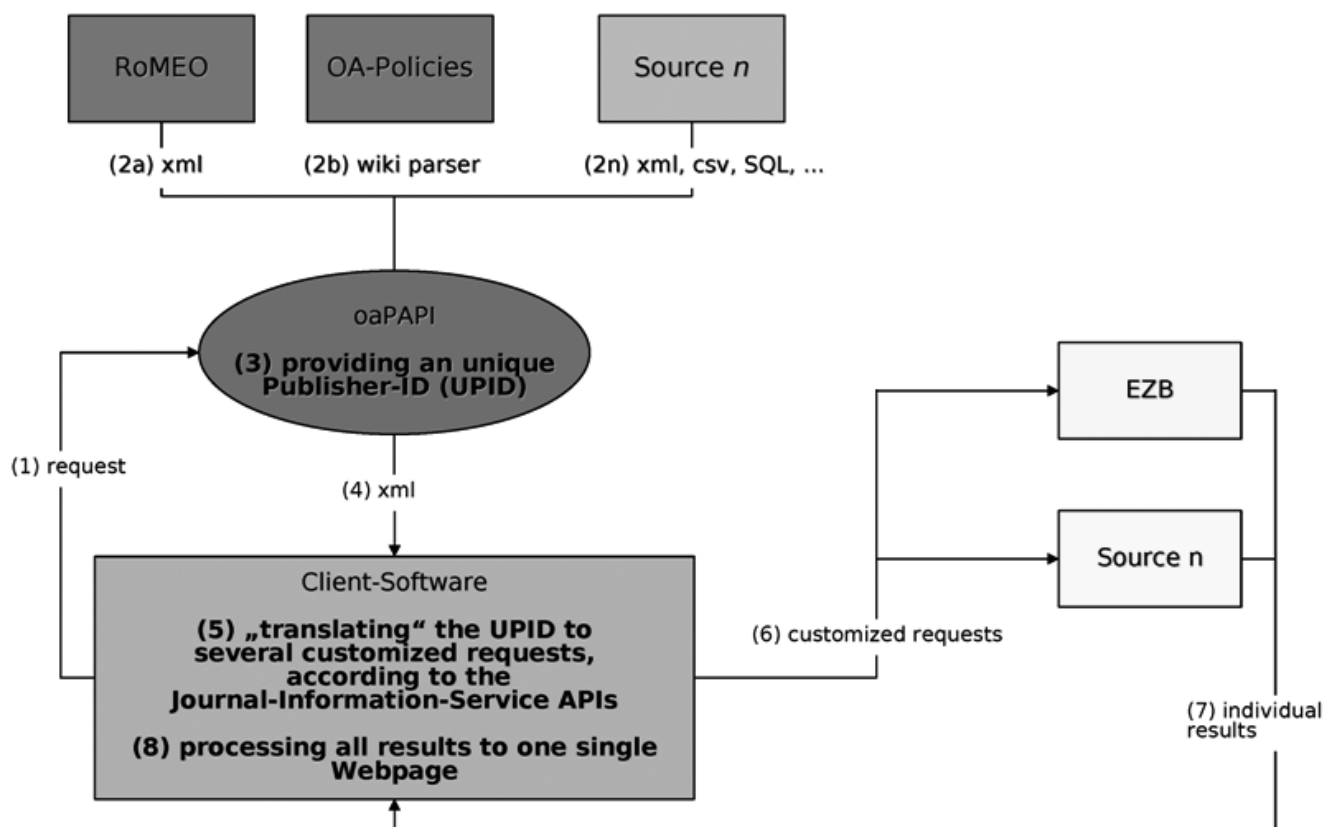


Abb. 3: Darstellung der Integration der EZB-Daten für den deutschsprachigen SHERPA/RoMEO-Client (oaPAPI.png)

MEO weiter auszubauen und eine Verknüpfung der Verlagsinformationen mit den Zeitschriftentiteln der Elektronischen Zeitschriftenbibliothek (EZB) [5] zu ergänzen.

Das Ziel dieser Verknüpfung ist zum einen die Kenntnis der Verlage über Open Access Policies zu befördern. Zum anderen dient sie der besseren Auffindbarkeit der Autorenrechte und stärkt damit die Position der Autoren gegenüber den Verlagen.

## Literatur

- [1] *Informationsplattform Open Access: Gründe für OA.* [http://open-access.net/de/allgemeines/gruende\\_und\\_vorbehalte/gruende\\_fuer\\_oa/](http://open-access.net/de/allgemeines/gruende_und_vorbehalte/gruende_fuer_oa/), letzter Zugriff: 27. Januar 2009
- [2] *ARBEITSGRUPPE ELEKTRONISCHES PUBLIZIEREN:* <http://edoc.hu-berlin.de/>, letzter Zugriff: 27. Januar 2009
- [3] *RoMEO Datenbank.* <http://www.sherpa.ac.uk/romeo.php>, letzter Zugriff: 27. Januar 2009
- [4] *deutschsprachiger SHERPA/RoMEO-Client.* <http://www.dini.de/wiss-publizieren/sherparomeo/>, letzter Zugriff: 27. Januar 2009
- [5] *ELEKTRONISCHE ZEITSCHRIFTENBIBLIOTHEK (EZB).* <http://rzblxi.uni-regensburg.de/ezeit/>, Letzter Zugriff: 27. Januar 2009

# Signatur und Zeitstempel zur Wahrung von Authentizität und Integrität

Niels Fromm | fromm@cms.hu-berlin.de

Unter den Begriffen Signatur oder elektronische Signatur und Zeitstempel versteht man Verfahren, die dazu dienen, Authentizität und Integrität von Daten in elektronischer bzw. digitaler Form zu sichern. Es können beliebige Arten von elektronischen Daten mit einer Signatur oder einem Zeitstempel versehen werden, z. B. auch Bild- oder Videodateien. Im Bereich des elektronischen Publizierens bestehen die elektronischen Daten aber meist aus Dokumenten wie MS Word- oder Adobe PDF-Dateien. Die wohl bekannteste Anwendung ist die Signatur von Rechnungen in elektronischer Form.

Authentizität bedeutet dabei, dass einem signierten Dokument eindeutig die Person zugeordnet werden kann, die die Signatur erstellt hat. Dies kann der Autor oder der Absender eines elektronischen Dokumentes sein. Unter Integrität wird in diesem Zusammenhang die Unversehrtheit z. B. eines elektronischen Dokumentes verstanden. Dies bedeutet, dass ein elektronisches Dokument vor Veränderungen oder Manipulationen geschützt ist. Die Fragen „Wer hat eine Nachricht gesendet?“ und „Wurde die Nachricht nach dem Absenden unbefugt verändert?“ sind schon bei den bisherigen Kommunikationsformen wie Telefon, Fax und Brief von großem Interesse. Sie gewinnen hinsichtlich von Dokumenten in elektronischer Form, die über das Internet übermittelt werden, zusätzlich an Bedeutung, da diese ohne besondere Sicherungen spurlos verändert und damit verfälscht werden können. Da elektronischen Dokumenten eine nachprüfbare Beziehung zu ihrem Urheber fehlt, sind im Internet Manipula-

tionen oder das Abstreiten von Handlungen möglich, ohne dass dies erkannt und verhindert werden könnte.

Eine rechtssichere Nutzung der elektronischen Kommunikation, wie der Kommunikation über E-Mail, ist allerdings nur denkbar, wenn solcher Missbrauch der technischen Möglichkeiten wirksam ausgeschlossen werden kann. Das Konzept der elektronischen Signaturen stellt eine Möglichkeit dar, eine im Sinne der Authentizität und Integrität von elektronischen Dokumenten sichere Kommunikation zu ermöglichen.

Nach der europäischen Signaturrichtlinie [1], die im Jahr 2001 durch das Signaturgesetz in deutsches Recht umgesetzt wurde, ist eine rechtssichere Kommunikation nur mit Hilfe von qualifizierten elektronischen Signaturen möglich. Zeitstempel sind eine besondere Form dieser Signaturen, die vorrangig der Feststellung eines bestimmten Zeitpunktes und nicht der Authentifizierung dienen. So kann mit einem Zeitstempel rechtsicher die Existenz von Daten zu einem bestimmten Zeitpunkt bewiesen werden, nicht aber, von wem diese Daten stammen.

*Der folgende Artikel beschreibt die Eigenschaften von Signaturen als Grundlage für die rechtssichere Nutzung elektronischer Kommunikation und deren Einsatz auf dem Dokumentenserver der Humboldt-Universität zu Berlin.*

*qualifizierte elektronische Signatur, Zeitstempel, Integrität, Authentizität*



Abb. 1: Signaturherstellungsgarät

## Trustcenter

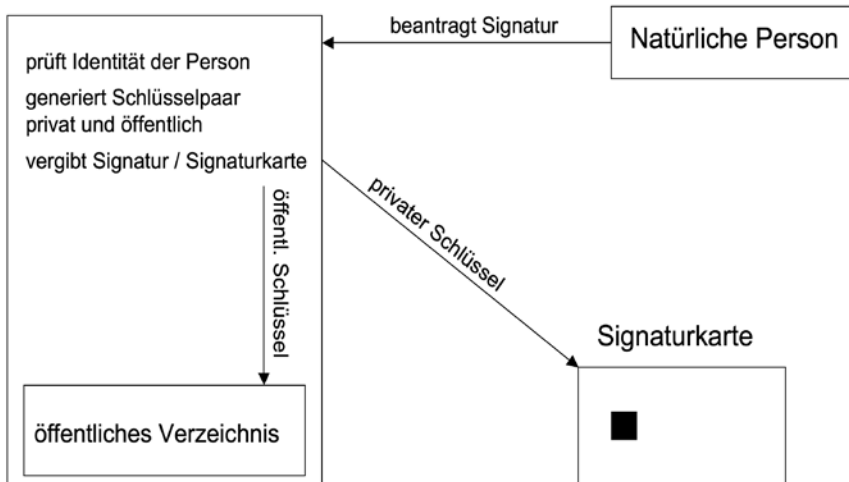


Abb. 1: Ablauf Ausgabe Signaturkarte

Technisch basiert eine qualifizierte elektronische Signatur auf zwei Verfahren: einem asymmetrischen Kryptografieverfahren und einer Hashfunktion. Bei asymmetrischen Kryptografieverfahren kommen zwei Schlüssel zum Einsatz. Es gibt einen öffentlichen und einen vom Nutzer geheim zu haltenden privaten Schlüssel, weshalb dieses Verfahren oft auch als „Public-Key“-Verfahren bezeichnet wird. Der öffentliche Schlüssel darf und soll veröffentlicht werden und kann dann von jedem anderen Anwender benutzt werden, um an den Eigentümer eine verschlüsselte Nachricht zu senden. Mit dem privaten Schlüssel des Eigentümers kann dieser dann die Nachricht wieder entschlüsseln. Bekannt ist die Anwendung eines solchen Public-Key-Verfahrens zum Beispiel bei der SSL-Verschlüsselung von Webseiten beim Online-Banking zur sicheren Eingabe von Daten. Beim Dokumentenserver wird diese Art der Verschlüsselung genutzt, um die Eingabe von persönlichen Daten bei der Abgabe von elektronischen Publikationen abzusichern.

Eine Hashfunktion ist im mathematischen Sinne eine Einwegfunktion, deren Ziel die Abbildung unterschiedlich großer Eingabemengen in eine kleine Ausgabemenge ist. Das Resultat einer solchen Hashfunktion bildet der sogenannte Hashwert. Somit können beliebig große Dateien durch einen kleinen Hashwert mit fester Größe eindeutig identifiziert werden. Kleinste Änderungen an der Datei resultieren in einem anderen Hashwert.

Zu einer qualifizierten elektronischen Signatur gehört auch ein sogenannter „vertrauenswürdiger Dritter“ oder Trustcenter, in Deutschland sind dies unter anderem die Trustcenter der Deutschen Telekom oder der Deutschen Post (siehe [2]). Für den Betrieb eines solchen Trustcenters gibt es strenge Regeln, deren Einhaltung das Bundesamt für Sicherheit und Informationstechnik (BSI) [3] überwacht. In der folgenden Abbildung wird der Ablauf der Ausstellung einer Signaturkarte für die Signaturerstellung und die handelnden Personen dargestellt.

Das Zertifikat in Form des öffentlichen Schlüssels ermöglicht die Identifizierung des Unterzeichners, da einer Signatur eine Person zugeordnet werden kann. Um eine Datei mit einer qualifizierten elektronischen Signatur zu ver-

sehen, benötigt man eine von einem beim BSI zertifizierten und akkreditierten Trustcenter ausgestellte Signaturkarte. Auch die für den gesamten Ablauf notwendige Software, sowie auch das benutzte Kartenlesegerät müssen vom BSI zertifiziert werden.

Beim Vorgang der Signaturerstellung wird zunächst von der zu signierenden Datei ein Hashwert erzeugt. Dieser Hashwert identifiziert die zu signierende Datei im Sinne des Signaturgesetzes eindeutig, wenn ein vom BSI zur Signaturerstellung erlaubtes Hashverfahren eingesetzt wird. Für den eigentlichen Vorgang der Signierung wird dieser Hashwert an die Signaturkarte übertragen, die den privaten Schlüssel des Karteneigentümers enthält.

Der private Schlüssel verlässt niemals die Signaturkarte, um einer möglichen Kompromittierung der Sicherheitseigenschaft vorzubeugen. Die eigentliche Unterschrift wird durch die mathematische Verknüpfung von Hashwert und privatem Schlüssel in der Signaturkarte erstellt und dann ausgegeben. Dieser Vorgang wird von der Signaturkarte aber nur nach erfolgreicher Authentifizierung des Inhabers durch Eingabe einer persönlichen Identifikationsnummer (PIN) durchgeführt.

Diese Kombination von Hashwert und Unterschrift wird als qualifizierte elektronische Signatur bezeichnet, wenn sie mit einer von einem akkreditierten Trustcenter herausgegebenen Signaturkarte erzeugt wurde. Eine durch eine qualifizierte elektronische Signatur geschützte Datei besteht meist aus der sig-

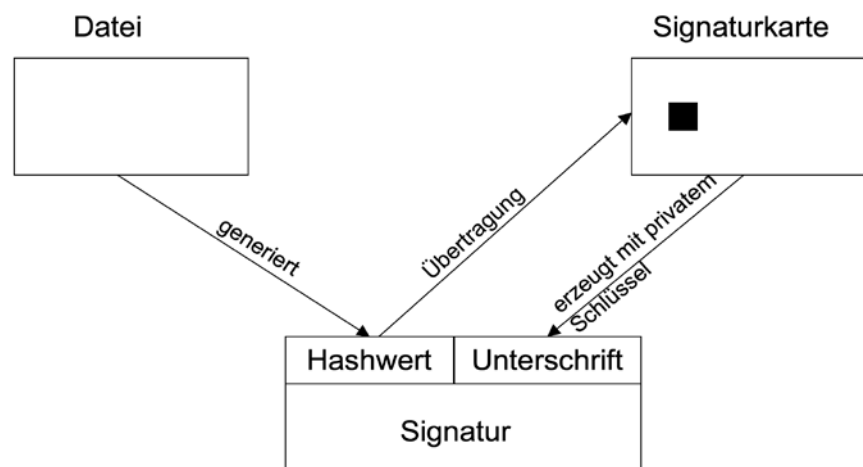


Abb. 2: Signieren einer Datei

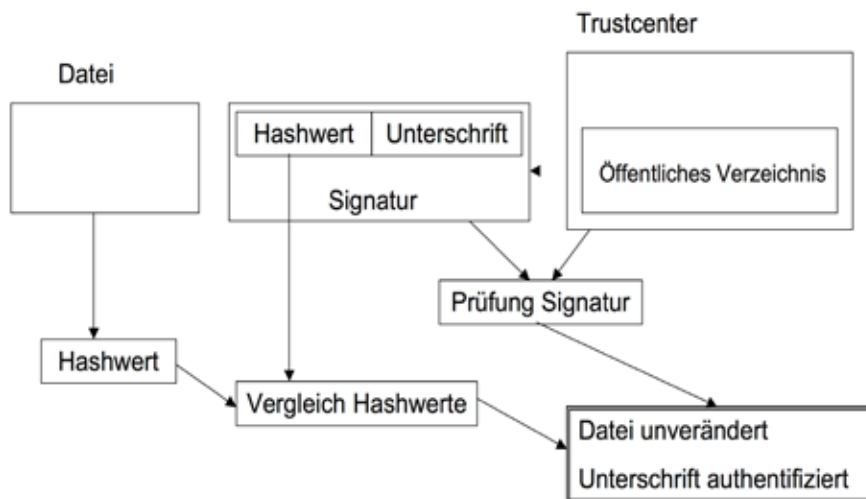


Abb. 3: Prüfen einer Signatur

nierten Datei selbst und einer weiteren Datei, die den Hashwert und die Unterschrift enthält.

Mit der Signatursoftware kann man auch ohne eine Signaturkarte die Signatur prüfen. Diese Prüfung erfolgt in zwei Stufen: Zunächst wird von der zu prüfenden Datei ein Hashwert mit dem gleichen Verfahren erzeugt, das für die Signatur verwendet wurde. Dieser Wert wird nun mit dem in der Signatur enthaltenen Wert verglichen. Stimmen beide überein, hat man den Nachweis, dass die zu prüfende Datei identisch mit der signierten ist und somit unverändert vorliegt.

Der zweite Teil der Prüfung besteht in der Prüfung der Unterschrift gegen das vorhandene Zertifikat, also dem öffentlichen Teil des Schlüsselpaars. Hier wird geprüft, von welchem Schlüsselpaar die Unterschrift erzeugt wurde bzw. welche Person diese Unterschrift geleistet hat. In dieser Kombination und unter Einbeziehung des Trustcenters lassen sich Integrität und Authentizität der Datei und qualifizierter Signatur rechtssicher beweisen. Für eine weitere Auseinandersetzung mit dem Thema sei das Buch „Grundlagen der elektronischen Signatur“ [4] empfohlen.

## Anwendung auf dem Dokumentenserver

Der Bestand des Dokumenten- und Publikationsservers der Humboldt-Universität zu Berlin soll archiviert und dabei

die Integrität der abgelegten Dokumente durch den Einsatz von Hashverfahren gesichert werden. Dazu wird für jede Datei auf dem Dokumentenserver ein Hashwert ermittelt, der mit dem auf dem Archiv- und Signaturserver gespeicherten Hashwert der entsprechenden Datei abgeglichen wird. Bei Übereinstimmung beider Hashwerte ist die Identität mit der archivierten Datei und somit deren Integrität gegeben. Durch einen automatisierten Vergleich der Hashwerte für alle auf dem Dokumentenserver vorhandenen Dateien kann die Integrität des gesamten Bestandes überprüft werden.

Die an den Dokumenten vorgenommenen Änderungen sollen erfasst und dokumentiert werden. Dies bedeutet, dass festgehalten wird, welcher Mitarbeiter aus welchem Grund eine Änderung

durchgeführt hat. Zusätzlich wird ein Teil der Dokumente, die universitären Qualifikationsarbeiten wie Dissertationen und Habilitationsschriften, durch elektronische Signaturen und Zeitstempel rechtssicher abgelegt werden. Damit wird die Möglichkeit geschaffen, den Zeitpunkt der elektronischen Veröffentlichung einer Qualifikationsarbeit und deren Unversehrtheit auch nach vielen Jahren vor Gericht beweisen zu können.

Voraussetzung für eine elektronische Veröffentlichung einer Qualifikationsarbeit an der Humboldt-Universität zu Berlin ist die Annahme dieser Arbeit durch die Arbeitsgruppe Elektronisches Publizieren. Dazu wird geprüft, ob die abgegebene Arbeit die von der Arbeitsgruppe aufgestellten Bedingungen zur Veröffentlichung von Dokumenten auf dem Dokumentenserver erfüllt. Nach der erfolgreichen Prüfung der Publikation wird sie an die Universitätsbibliothek weitergeleitet, die daraufhin die Publikationsbescheinigung ausstellen kann. Gleichzeitig wird die Publikation in dem Publikationsformat PDF dem Bestand des Dokumentenservers der Humboldt-Universität zu Berlin hinzugefügt. Damit ist die Publikation über die Webseiten des Dokumentenservers im Internet frei zugänglich und gilt somit als veröffentlicht.

Dies ist daher der Zeitpunkt der elektronischen Veröffentlichung einer Publikation, der rechtssicher dokumentiert werden soll. Dazu wird zu diesem Zeitpunkt über die gesamte Publikation ein Zeitstempel erzeugt.

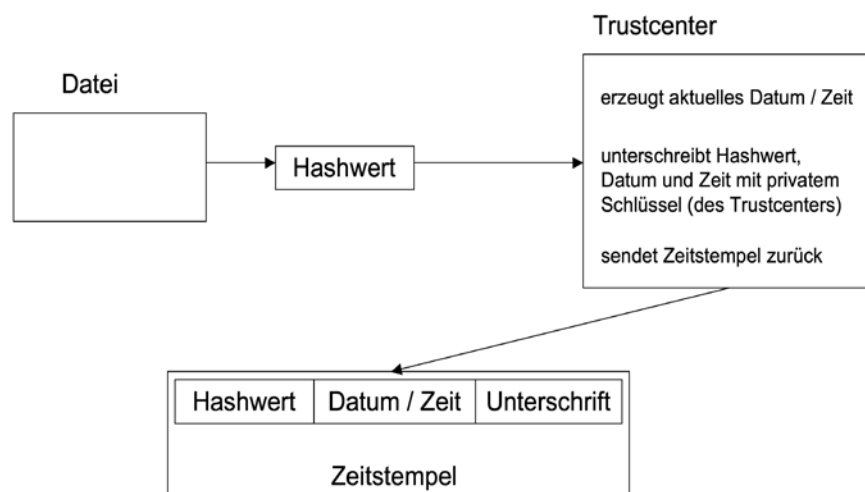


Abb. 4: Erzeugung eines Zeitstempels

Wie bei der Signierung von Dateien wird zunächst wieder ein Hashwert der Datei erzeugt, die einen Zeitstempel erhalten soll. Im Anschluss daran wird dieser Hashwert an ein Trustcenter geschickt. Das Trustcenter ermittelt die Zeit und das Datum bei Erhalt dieses Hashwertes. Dann unterschreibt das Trustcenter den Hashwert und die ermittelte Zeit und Datum, analog zur Erstellung einer Signatur, jedoch mit Signaturkarten des Trustcenters. Die drei Elemente Hashwert, Zeit bei Erhalt des Hashwertes und die Unterschrift des Trustcenters bilden zusammen den sogenannten Zeitstempel. Das Trustcenter bezeugt, dass der vorgelegte Hashwert zu der Zeit der Annahme existiert hat. Da der Hashwert eine bestimmte Datei identifiziert, beglaubigt das Trustcenter also indirekt, dass eine bestimmte Datei zu einem bestimmten Zeitpunkt vorgelegen hat. Mit diesem Zeitstempel lässt sich dann rechtssicher beweisen, dass diese Publikation zu dem im Zeitstempel festgehaltenen Zeitpunkt in genau dieser Form existiert hat.

Weiter wird diese Art von Publikation bei jeder Änderung mit einer qualifizierten elektronischen Signatur versehen und somit festgehalten, welcher Mitarbeiter diese Änderungen vorgenommen hat. Alle Publikationen und zugehörigen Hashwerte, Zeitstempel und Signaturen werden auf einem gesonderten Archiv- und Signaturserver getrennt vom eigentlichen Dokumentenserver archiviert.

Dieser Server wird besonders gesichert und ist nicht über das Internet öffentlich zugänglich. Zugriff haben nur Mitarbeiter der Arbeitsgruppe von bestimmten Arbeitsplätzen aus. Der Archiv- und Signaturserver muss derart gesichert werden, denn die darauf gespeicherten Hashwerte, Zeitstempel und Signaturen bilden den Beweis für die Integrität und Authentizität aller auf dem Dokumentenserver der Humboldt-Universität zu Berlin angebotenen Publikationen.

## Rechtssichere Archivierung

Bei der rechtssicheren Archivierung sollen Dokumente in elektronischer Form für einen längeren Zeitraum, in der Regel

mindestens zehn Jahre, aufbewahrt und dabei die Gültigkeit ihrer elektronischen Signaturen erhalten werden.

Eine qualifizierte elektronische Signatur ist unbegrenzt gültig, wenn sie nach den rechtlichen Vorgaben erstellt wurde und das für die Signatur genutzte Zertifikat zum Zeitpunkt der Signaturerzeugung nicht gesperrt war. Nach längerer Zeit der Archivierung entstehen jedoch Probleme beim Versuch zu beweisen, dass ein Zertifikat bei der Erzeugung nicht gesperrt war und die damit erzeugte Signatur somit ungültig ist. Noch wichtiger ist der Beweis, dass die signierten Dokumente wirklich noch genau den ursprünglich signierten entsprechen oder ob sie verfälscht wurden.

In der Signaturverordnung ist daher ein Verfahren zur langfristigen Sicherung von signierten Dokumenten festgelegt. Dabei sind die mit einer qualifizierten elektronischen Signatur versehenen Dokumente „(...) vor dem Zeitpunkt des Ablaufs der Eignung der Algorithmen oder der zugehörigen Parameter mit einer neuen qualifizierten elektronischen Signatur zu versehen. Diese muss mit geeigneten neuen Algorithmen oder zugehörigen Parametern erfolgen, frühere Signaturen einschließen und einen qualifizierten Zeitstempel tragen.“[5] Dieses Verfahren wird als Neu- oder Übersignatur bezeichnet.

Das Konzept der Übersignatur ist somit die Grundlage für Projekte im Bereich der rechtssicheren Archivierung. Das Pilotprojekt in Deutschland für den Einsatz von elektronischen Signaturen in der Archivierung von Dokumenten ist das Projekt Archisafe [6] der Physikalisch-Technischen Bundesanstalt (PTB) [7]. Der Fokus des Projektes lag vor allem darauf, die Übersignatur für eine große Anzahl von Dokumenten effizient durchführen zu können.

Für die Archivierung der Dokumente und Signaturen wurde jedoch ein kommerzielles Produkt eingesetzt, daher ist eine Nachnutzung der Ergebnisse dieses Projektes nur bedingt möglich. Auch sind die Dokumente dabei ausschließlich PDF/A-Dokumente mit integrierter Signatur. Dabei soll die langfristige Lesbarkeit der Dokumente allein durch Verwendung des PDF/A-Dokumentformats erreicht werden.

Im Gegensatz dazu soll an der Humboldt-Universität zu Berlin die langfristige Lesbarkeit durch die kontinuierliche Migration der Dokumente sichergestellt werden. Für die rechtssichere Archivierung von Qualifikationsarbeiten der Humboldt-Universität zu Berlin wurde daher ein eigenes Konzept entwickelt [8].

Das ursprüngliche Konzept für die rechtssichere Archivierung wurde vom Autor weiterentwickelt und für eine genaue Beschreibung wird auf „Einsatz elektronischer Signaturen auf dem Dokumentenserver der Humboldt-Universität zu Berlin“ [9] verwiesen.

## Literatur

- [1] *Richtlinie 1999/93/EG des Europäischen Parlaments und des Rates vom 13. Dezember 1999 über gemeinschaftliche Rahmenbedingungen für elektronische Signaturen.* 2001
- [2] *T7 e.V. – Arbeitsgemeinschaft der Trustcenterbetreiber.* URL: <http://www.t7ev.org/>
- [3] *Bundesamt für Sicherheit in der Informationstechnik.* URL: <http://www.bsi.de/>
- [4] *Hühnlein & Korte: Grundlagen der elektronischen Signatur.* 2006
- [5] *Bundesministerium für Justiz: Verordnung zur elektronischen Signatur.* 2001 d, S. 3079 §17.
- [6] *Archisafe.* URL: <http://www.archisafe.de>
- [7] *Physikalisch-Technische Bundesanstalt.* URL: <http://www.ptb.de> (letzter Zugriff am 04.12.2008)
- [8] *Ohst: Einsatz elektronischer Signaturen und Zeitstempel für die Sicherung digitaler Dokumente.* 2004
- [9] *Fromm: Einsatz elektronischer Signaturen auf dem Dokumentenserver der Humboldt-Universität zu Berlin.* 2008



# Backup-Strategie für den Dokumentenserver

Niels Fromm | fromm@cms.hu-berlin.de

Ein wichtiger Teil des Konzeptes für die Sicherheit der auf dem Dokumentenserver abgelegten Daten ist eine Backup-Strategie, die diese Daten vor Verlust oder Beschädigung schützt. Ein Verlust von Daten oder deren Beschädigung kann durch fehlerhafte Hard- oder Software, unbeabsichtigte menschliche Fehler, aber auch durch gezielte Manipulation, etwa durch Schadprogramme wie Viren oder Würmer, entstehen. Um in einem solchen Fall die verlorenen oder beschädigten Dateien wiederherstellen zu können, muss es eine Kopie dieser Dateien, ein Backup, geben.

für den Dokumentenserver. Der Computer- und Medienservice (CMS) bietet daher für die gesamte Universität einen Backup-Service [1] an. Dieser Service bietet die Möglichkeit, alle Daten eines Servers automatisch jede Nacht in den Backupspeicher der Universität zu übertragen. Dieser Speicher besteht aus zwei Magnetband-Speichertürmen mit rund 200 TB Speicherkapazität, wobei sich ein Speicherturm im Rechnerraum in Mitte, der andere im Erwin Schrödinger-Zentrum in Adlershof befindet.

Der Inhalt beider Speichertürme wird jede Nacht synchronisiert, so dass

*Der folgende Artikel bietet eine Übersicht über die Backup-Strategie der Humboldt-Universität zu Berlin und beschreibt den Ansatz der Arbeitsgruppe Elektronisches Publizieren für die Sicherung der Daten auf dem Dokumentenserver.*

*Backup, Redundanz, SAN*



Abb. 1: Backup-Speicherturm des CMS im Erwin Schrödinger-Zentrum

Die Notwendigkeit von einem Backup der Daten besteht für alle Server der Humboldt-Universität zu Berlin, nicht nur

zu jedem Zeitpunkt alle Daten sowohl in Mitte als auch in Adlershof komplett vorhanden sind. Dadurch ist sichergestellt,

dass alle Daten selbst bei komplettem Ausfall eines Speicherturmes noch verfügbar sind.

Aus diesem Backupspeicher lassen sich beschädigte oder verlorene Dateien der letzten 60 Tage wiederherstellen. Der Backup-Service des CMS bietet seinen Nutzern zudem eine Versionierung, d. h. die vier letzten Versionen einer Datei aus diesem Zeitraum können zurückkopiert werden. Daten von Servern, die den Backup-Service nutzen, sind also dreifach redundant gespeichert, einmal auf dem lokalen Speicher des Servers selbst und zusätzlich zweimal auf den zentralen Backup-Speichertürmen.

Das beschriebene Backup der Daten von Servern auf einen zentralen Backupspeicher auf Magnetbandbasis ist die traditionelle, aber nach wie vor aktuelle und praktisch überall in Rechenzentren eingesetzte Methode, da nur sie die relativ kostengünstige Speicherung von sehr großen Datenmengen ermöglicht.

Ein Nachteil der Speicherung des Backups auf Magnetbändern ist die im Vergleich relativ geringe Geschwindigkeit der Medien. So kann das Wiederherstellen von größeren Datenmengen einige Zeit dauern.

Für das schnelle Wiederherstellen von Servern im Havariefall, also z. B. bei Ausfall von Komponenten des Servers, ist diese Art des Backups nicht unbedingt geeignet. Für Server wie den Dokumentenserver, die nach Möglichkeit jederzeit ihre Dienste anbieten sollen, müssen zusätzliche Sicherungsmaßnahmen getroffen werden. In Bezug auf die Sicherheit der Daten ist dafür vor allem ein ausfallsicheres, hochverfügbares Speichersystem für die Server nötig.

Der CMS hat aus diesem Grund vor einiger Zeit ein sogenanntes Storage Area Network (SAN) aufgebaut (siehe [2], [3]). Es besteht aus Festplattenspeichersystemen, die mit den Servern über ein eigenes Netz verbunden sind. Die Verbindung eines solchen Speichernetzwerkes wird typischerweise über Lichtwellenleiter (LWL), also Glasfaserkabeln hergestellt, sie ist damit unabhängig von der Anbindung des Servers an das normale Netzwerk. Ein an ein SAN angeschlossener Server kann also Speicher nutzen, der an be-

liebiger Stelle in diesem Netz angeboten wird, der Speicher muss nicht mehr in dem Server selbst vorhanden sein.

Dies ermöglicht eine zentrale Administration und vor allem Planung und damit effiziente Nutzung der gesamten Festplattenkapazitäten. In vielen Fällen ist dieser Speicher auch leistungsfähiger als lokal in Server eingebaute Festplatten.

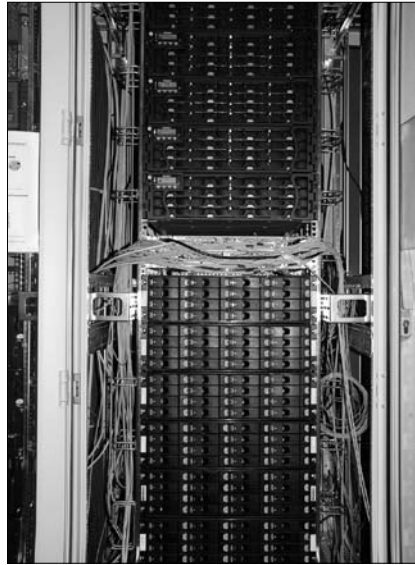


Abb. 2: SAN-Festplattensysteme im Erwin Schrödinger-Zentrum

Das SAN der Humboldt-Universität zu Berlin ist vor allem unter dem Gesichtspunkt der Ausfallsicherheit entwickelt worden. Das Konzept des CMS für die Benutzung des SAN sieht Redundanz auf jeder Ebene vor, damit der Ausfall einer Komponente nicht den Ausfall des gesamten Systems nach sich zieht. So muss zunächst jeder Server zweifach an das SAN angeschlossen werden und auch die gesamte Verbindung zwischen den Servern und den Speichersystemen ist redundant ausgelegt. Damit ist der Speicher für den Server auch dann noch verfügbar, wenn ein Anschluss an das SAN gestört ist.

Der Speicher im SAN besteht aus redundanten Festplattensystemen, die auf verschiedene Gebäude der Universität verteilt sind und den Ausfall einer einzelnen Festplatte ohne Verlust von Daten überstehen können. Als zusätzliche Sicherheit werden jeweils zwei dieser Festplattensysteme gespiegelt und den Servern als ein Speichersystem angeboten. Alle Daten werden daher gleichzeitig auf zwei der redundanten Speichersysteme geschrieben, die sich auch an unterschiedlichen Orten befinden können. Somit kann auch ein gesamtes Speichersystem ausfallen, ohne die Sicherheit der Daten zu gefährden. Weiter bieten die Speichersysteme im SAN zusätzliche Möglich-

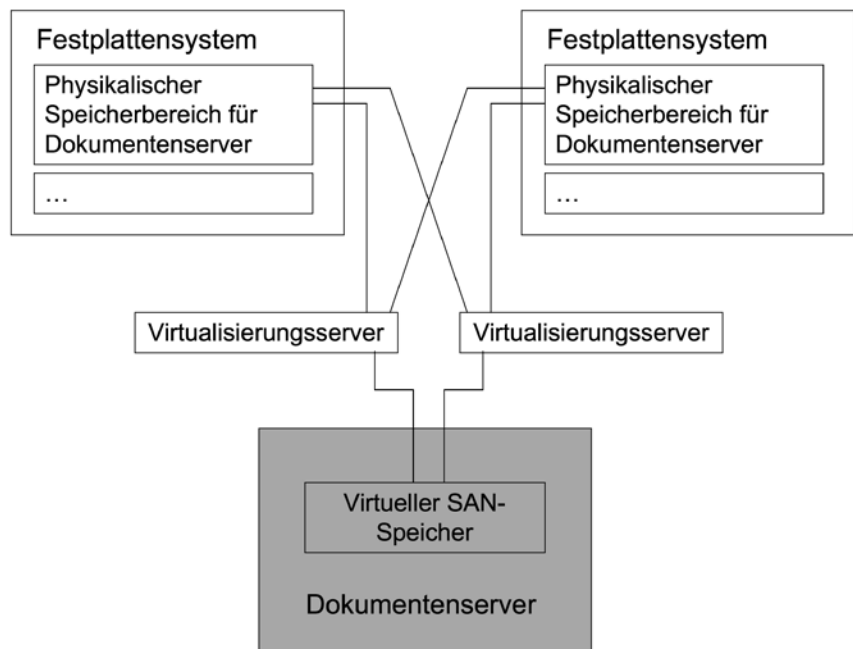


Abb. 3: Anbindung Dokumentenserver – SAN

keiten der Datensicherung wie Snapshots, die bei der Backup-Strategie für den Dokumentenserver Anwendung finden.

Der Dokumentenserver der Humboldt-Universität zu Berlin nutzt zurzeit zwei virtuelle Festplatten mit je 400 GB Kapazität aus dem SAN. Der Dokumentenserver befindet sich im Rechnerraum des CMS im Erwin Schrödinger-Zentrum in Adlershof. Die 800 GB der beiden virtuellen Festplatten befinden sich physikalisch in Festplattensystemen in den Rechnerräumen des CMS in Adlershof und im Hauptgebäude der Universität in Mitte. Jeden Monat einmal werden diese 800 GB des Dokumentenservers als zusätzliches Backup auf andere Speichersysteme im SAN der Universität abgelegt. Dieses zusätzliche Backup liegt in Speichersystemen im Erwin Schrödinger-Zentrum und im Gebäude der Physik in Adlershof. Weiterhin wird jeden Tag ein sogenannter Snapshot der Daten des Dokumentenservers im SAN angelegt. Ein Snapshot von einem Dateisystem kann zu jedem beliebigen Zeitpunkt erstellt werden. Jede noch so kleine Änderung an Daten in diesem Dateisystem nach diesem Zeitpunkt wird im Snapshot festgehalten und kann dadurch rückgängig gemacht werden. Da in einem Snapshot alle Änderungen an den Daten seit seiner Erstellung gespeichert werden, vergrößert sich ein Snapshot mit jeder Änderung, er wächst also mit der Zeit. Je mehr Daten sich in diesem Dateisystem verändern, desto schneller wächst die Größe des Snapshots.

Um nicht zuviel Speicherplatz zu verwenden, wird die Größe aller Snapshots für den Dokumentenserver auf ein TB begrenzt und bei Erreichen dieser Grenze werden automatisch die ältesten Snapshots gelöscht bzw. überschrieben. Der Speicherplatz für diese Snapshots befindet sich physikalisch in Speichersystemen im ESZ und im Gebäude der Informatik und reicht bei täglichen Snapshots für aktuell 180 Tage. Alle Änderungen an den im Dateisystem abgelegten Daten des Dokumentenservers der letzten 180 Tage können also rückgängig gemacht werden, also z. B. auch gelöschte oder beschädigte Dateien wiederhergestellt werden.

Selbstverständlich nutzt der Dokumentenserver auch den Backup-Service

des CMS, jede Nacht werden daher die Daten des Dokumentenservers auf den beiden Backup-Speichersystemen der Universität gesichert.

Die Backup-Strategie für den Dokumentenserver legt also fest:

- die Nutzung von Speicher im SAN und damit die redundante Speicherung dieser Daten im Erwin Schrödinger-Zentrum und in Mitte,
- die tägliche Erstellung von Snapshots von diesen Daten und die redundante Speicherung der Snapshots im Erwin Schrödinger-Zentrum und in der Informatik,
- das tägliche Backup der Daten in die redundanten Backup-Speichersysteme im Erwin Schrödinger-Zentrum und in Mitte
- und die tägliche Kopie aller Daten sowie deren redundante Speicherung in Mitte und der Physik.

Damit sind alle Daten des Dokumentenservers nach physikalisch sechs Mal an vier verschiedenen Orten Berlins vorhanden, was einen sehr hohen Grad an Sicherheit für diese Daten bedeutet. Denn dass alle acht Kopien der Daten an den verschiedenen Standorten gleichzeitig verloren gehen oder zerstört werden, kann als nahezu unmöglich bezeichnet werden.

Die Anzahl der vorhandenen Kopien von Daten wird demnach als Maß für die Sicherheit dieser Daten vor Verlust oder Beschädigung gesehen. Aber die räumliche Trennung der verschiedenen Kopien ist ebenso wichtig wie die Anzahl der Kopien, denn existieren alle Kopien an einem Ort, so können sie auch gleichzeitig zerstört werden.

Bei der Erstellung der Backup-Strategie für den Dokumentenserver wurde daher versucht, eine hohe Anzahl von Kopien der Daten auf möglichst viele unterschiedliche Standorte zu verteilen, um ein Höchstmaß an Sicherheit zu erreichen. Dies ist mit der vorliegenden Strategie gelungen, wobei dies nur möglich ist, da die Humboldt-Universität zu Berlin mit dem SAN über die nötige technische Infrastruktur verfügt und die Datenmenge mit weniger als 800 GB noch so klein ist, dass die Erstellung von mehreren Kopien kein Problem darstellt.

## Literatur

- [1] WEICKMANN, CHRISTOPH: *Backup – Datensicherheit für alle*. cms-journal Nr. 25, 2004, System- und Netzdienste S. 63–64
- [2] SITTEL, FRANK: *Storage-Area-Network an der Humboldt-Universität*. cms-journal Nr. 22, 2001, 10 Jahre RZ-Mitteilungen S. 10–14
- [3] SITTEL, FRANK: *Institute ans SAN*. cms-journal Nr. 25, 2004, System- und Netzdienste S. 59–61

# Open Access – Verfügbar ist noch nicht präsent

## Verbesserung der Sichtbarkeit von Open Access Repositories durch die Bildung von Netzwerken

Robin Malitz | malitzro@cms.hu-berlin.de

### Suchmaschinen und das „Unsichtbare Netz“

Um der Fragestellung der Sichtbarkeit von Open-Access-Dokumenten nachzuspüren, beginnt man am besten aus der Sicht des Suchenden, der sich ins Netz begibt, um nach inhaltlich für ihn relevanten, wissenschaftlichen Volltexten zu recherchieren. Der erste Anlaufpunkt sind im einfachsten Falle die gängigen Suchmaschinen, die man in der Hoffnung mit Schlagworten füttert, dass sie das, was im Netz verfügbar ist, auch finden. Schnell wird jedoch klar: Das gezielte, fachspezifische Recherchieren nach online publizierten Dokumenten gerät zur mühseligen Goldwäscherei in den angezeigten Trefferseiten. Selbst vielversprechende Links führen nicht immer zum gewünschten Volltext.

Besser scheint da, wird direkt der Recherchezugang eines Open Access Repositories genutzt. Auch hier gibt es heutzutage in fast allen Standardlösungen an zentraler Stelle eine Suchmaske. Auf Knopfdruck erhält man eine Trefferliste, die ausschließlich aus wissenschaftlichen Dokumenten besteht, deren Volltexte man einsehen, herunterladen und ausdrucken kann. Mehr noch, häufig kann man viel detaillierter in einer Art Expertensuche gezielt nach Veröffentlichungen mit bestimmten Eigenschaften recherchieren. Hier unterscheiden sich die angebotenen Suchmöglichkeiten von Repository zu Repository. Die Vorteile dieser Recherchemöglichkeit liegen jedoch auf der Hand.

Eine Kehrseite gibt es auch hier. In der Regel kann nur die lokal archivierte Menge an Dokumenten gezielt durch-

sucht werden. Was ist jedoch mit Dokumenten, die an anderen Einrichtungen publiziert worden und nicht in dem gerade durchsuchten Repository verfügbar sind? Diese lassen sich im Letzteren dann nicht finden. Da es allein in Deutschland über einhundert wissenschaftliche Open Access Repositories gibt – DINI listet 138 verschiedene Webadressen auf – ist eine lokale Suche bei einzelnen Repositories weder sinnvoll noch komfortabel.

Wer sich nun geschlagen gibt und wieder die Suchmaschinen bemüht, dem fällt vielleicht sogar auf, dass einige in Repositories frei verfügbare Dokumente nicht gefunden werden. Das liegt daran, dass die Rechercheoberflächen der Repositories dynamische Webseiten sind, die die nimmermüden, automatischen Späher der Suchmaschinen, die Robots und Webcrawler, technologisch durch Software-Barrieren aussperren. Für die Suchmaschinen sind die archivierten Dokumente dadurch zumeist unsichtbar, weshalb dieses Phänomen als „Invisible Web“ oder „Deep Web“ bezeichnet wird. Auch wenn die Suchmaschinen immer leistungsfähiger werden, müssen sich die Betreiber von Repositories häufig Umwege ausdenken, um wenigstens die Volltexte der Dokumente – wenn auch nicht die dazugehörigen, für Recherchierende so wertvollen Metadaten – für die Suchmaschinen-Welt sichtbar zu machen. Verfügbar ist eben noch nicht präsent.

*Der Open-Access-Gedanke propagiert den freien Zugriff auf wissenschaftliche Texte – lesen und gelesen werden. Es steht außer Frage, dass das Veröffentlichen elektronischer Dokumente in Open Access Repositories diesen freien Zugriff gewährt. Doch genügt dieses, um gelesen zu werden? Sicher nicht. Impact kann nur haben, wer auch wahrgenommen wird. Sichtbarkeit heißt die wesentliche Herausforderung für Open Access und Vernetzung fördert diese auf verschiedene Weise.*

*Der Artikel skizziert die Hintergründe und grundlegenden Zusammenhänge, die sich aus dem Wechselspiel von Open-Archive-Welt und den etablierten Mechanismen des World Wide Web ergeben, und stellt u. a. die Bedeutung des auf die deutsche Repositories-Landschaft fokussierten Projektes „Open-Access-Netzwerk“ im Kontext der Sichtbarkeit dar.*

*Sichtbarkeit, Spezialsuchdienst, Open-Access-Netzwerk, Service Provider, Infrastruktur, Informationsintegration, Mehrwertdienste, Webservice*

## Von Spezialsuchdiensten zu Rechercheplattformen

Die Problematik ist bekannt: Schon früh erkannte man die Notwendigkeit, dass die in den Repositories gesammelten Publikationen Spezialsuchdiensten verfügbar gemacht werden müssen. Diese Spezialsuchdienste sollen für den Suchenden die Grenzen zwischen den Repositories verwischen und ihm dabei erlauben, möglichst viele der zu einem Dokument erfassten Metadaten in seine Suchanfrage einzubeziehen. Wird z. B. nach Dokumenten eines bestimmten Autors zu einem bestimmten Erscheinungsjahr gesucht, dann soll der Spezialsuchdienst alle passenden Dokumente als Treffer präsentieren, unabhängig davon, wo das Dokument tatsächlich archiviert worden ist. Solche Spezialsuchdienste gibt es bereits. Im Open Access Bereich bieten bislang Webangebote wie OAIster, Google Scholar, BASE, DRIVER und andere den Suchenden Recherche-Möglichkeiten.

Der Begriff „Spezialsuchdienst“ ist bei den aktuellen Projekten und Angeboten eigentlich zu einseitig. Recherche durch die Option einer detaillierten Suchanfrage ist nur ein Mehrwert-Dienst von vielen, den der Recherchierende im Kontext wissenschaftlicher Publikationen benötigt. Hinzu kommen Nachweis-Dienste (Browsing), Alerting, Newsfeeds, Export von Metadaten in Literaturverwaltungsformate und viele mehr.

Heutzutage spricht man daher oft von Portalen oder Plattformen. Der erste Begriff betont, dass der Nutzerzugang zu ganz verschiedenen Mehrwertdiensten gebündelt angeboten wird, der zweite, dass das Angebot technologisch so gestaltet wurde, dass das bestehende System als Basis und Präsentationsoberfläche für weitere Mehrwertdienste genutzt werden kann.

## Infrastruktur für hochwertige Mehrwertdienste

Repositoriesbetreiber, auf der anderen Seite, sind nun bestrebt, ihre Dokumente in Plattformen mit möglichst vielfältigen und hochwertigen Mehrwertdiensten an-

zubieten. Autoren publizieren vornehmlich dort, wo sie glauben, gelesen zu werden, und gelesen wird im Netz, wo man komfortabel recherchieren kann. Die Güte von Mehrwert-Diensten hängt unmittelbar vom Zusammenspiel von Plattformen und Repositories ab, von der Infrastruktur, die man dadurch erschafft. Daher lohnt es sich, zunächst einen Blick auf die technologischen Hintergründe zu werfen – sowohl auf Seiten der Plattformen als auch auf Seiten der Repositories.

Konzeptuell haben sich für die beiden Blickwinkel in der Fachwelt zwei Begriffe etabliert: Die Repositories, die in der Rolle als Datenlieferanten auftreten, werden als Data Provider bezeichnet. Das Gegenstück, die Plattformen, auf denen man Mehrwertdienste anbieten kann, stellen die Service Provider dar. Im Folgenden werden beide Begriffspaare analog benutzt.

Ganz allgemein betrachtet, müsste eine Plattform, also ein Service Provider, zunächst Zugriff auf die Daten aller Repositories haben, die für die Mehrwertdienste der Plattform als Data Provider auftreten sollen. Beginnt ein Service Provider erst nach der Suchanfrage eines Nutzers damit, die Daten der verschiedenen Repositories abzufragen und auszuwerten (ähnlich wie Cross-Search-Verfahren), würde eine Antwort des Dienstes an den Nutzer bei vielen teilnehmenden Repositories sehr lange dauern. Alternativ den Service Provider nur auf Basis eines einzigen, gigantischen Repositories anzubieten, das Kopien aller Daten und

aller Volltext-Dateien vorrätig hält und diese parallel zu den Ursprungs-Repositories pflegt, ist jedoch sehr ineffizient und in größeren Dimensionen nicht mehr sinnvoll zu verwalten. In der Realität wurde daher ein praktischer Kompromiss gefunden: Die Volltexte der Dokumente verbleiben physisch in den Repositories, wo sie bei nachhaltiger Planung langzeitarchiviert werden können. Nur die Metadaten – darunter Hyperlinks zu den Volltexten – werden an die Service Provider übertragen, wo sie als Kopien gespeichert und in einer gemeinsamen Datenbasis für die Mehrwertdienste vorgehalten werden.

Es bleibt also die Frage, nach welchem Mechanismus die Repositories den Service Providern ihre Metadaten anbieten können. Der seit etlichen Jahren im Produktiveinsatz befindliche und den Kennern der Thematik hinlänglich bekannte Standard sind die so genannten OAI-PMH-2.0-Schnittstellen. Dies sind spezielle, maschinenlesbare Schnittstellen, auf die die Software der Service Provider automatisiert zugreifen kann, um sich Metadaten in verschiedenen XML-Formaten abzuholen. Um die Daten aktuell zu halten, ohne immer den gesamten Datenbestand kopieren zu müssen, sind die Schnittstellen so konstruiert, dass die abfragenden Programme im Normalfall nur Änderungen bezüglich des letzten Durchlaufs beziehen können. Diesen Mechanismus des inkrementellen Updates nennt man daher auch „Harvesting“, zu Deutsch „Ernte“.

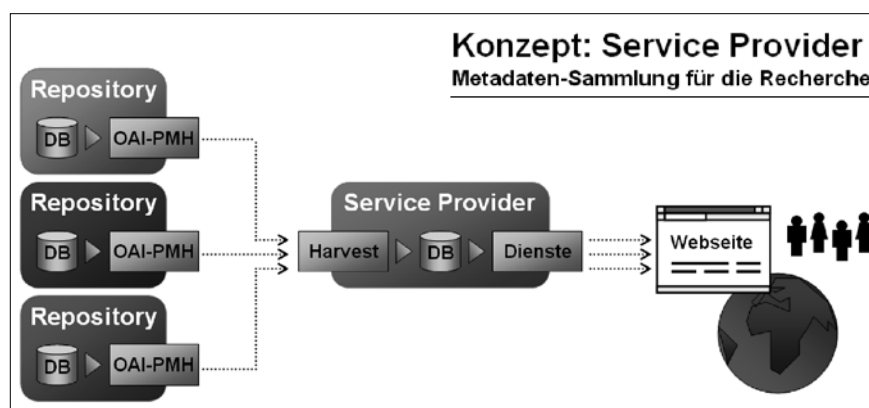


Abb. 1: Mittels des OAI-PMH 2.0 Harvesting-Mechanismus werden Metadaten in der Datenbank eines Service Providers hinterlegt, welcher daraufhin Dienste in einer zur Recherche bestimmten Webseite einer weltweiten Leserschaft anbieten kann.



## Herausforderungen und Chancen eines Netzwerks

Die Landschaft der Open Access Repositories ist noch immer sehr heterogen. Im August 2007 startete daher das von der Deutschen Forschungsgemeinschaft geförderte Projekt „Open-Access-Netzwerk“. Dessen Ziel ist es, die Betreiber von deutschen Open Access Repositories zusammenzubringen und zu unterstützen.

sollen so homogen zu einem gemeinsamen, virtuellen Datenraum zusammengefügt werden, dass der Nutzer der darauf angebotenen Mehrwert-Dienste deren ursprüngliche Verteiltheit nicht mehr wahrnimmt. Dazu müssen die eingesammelten Metadaten fehlerkorrigiert, harmonisiert, in Bezug zueinander gesetzt und mit weiteren, aus dem Kontext der akkumulierten Datenbasis bezogenen Daten angereichert werden.

dem Nutzer direkt neue Sichten und Mehrwerte bieten oder die Datenqualität des Systems weiter erhöhen, wird die API für die REST-Schnittstelle offengelegt. Das Partnerprojekt „Open-Access-Statistik“ wird auf diese Art und Weise Nutzungsstatistiken einbringen und unter anderem auf der gemeinsamen, webbasierten Benutzeroberfläche anbieten. Weitere Projekte sind in Planung.

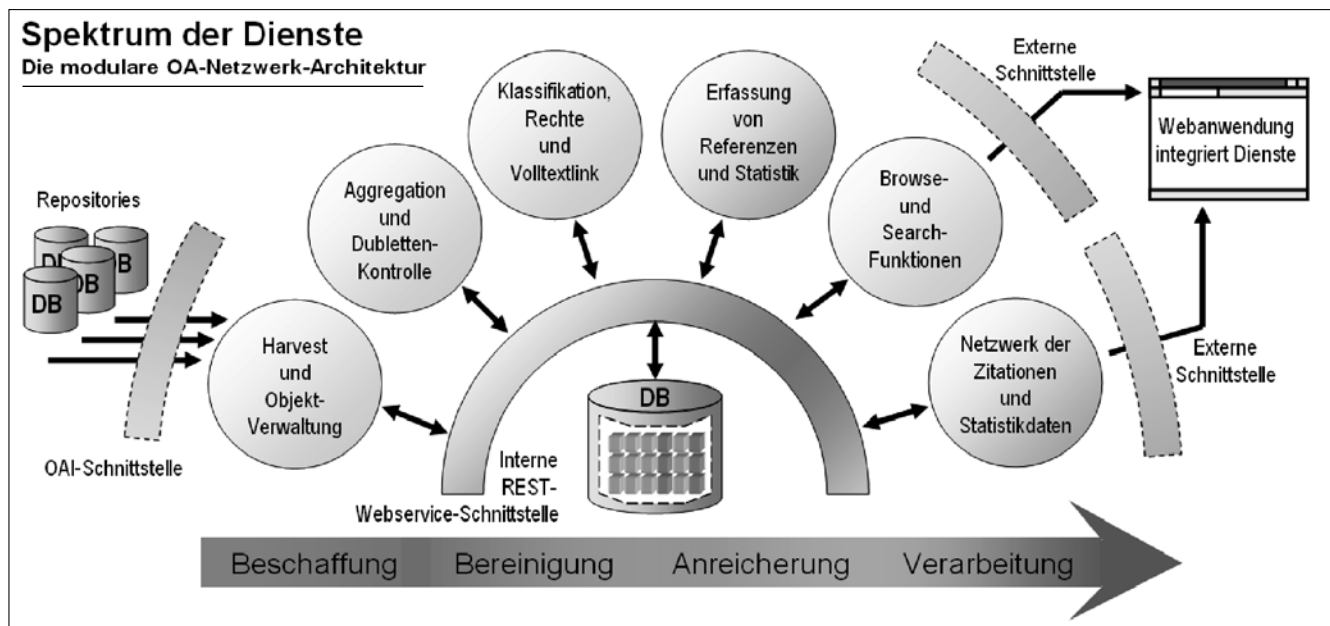


Abb. 2: Die Architektur des Service Providers von Open-Access-Netzwerk ist aus verteilten arbeitenden Modulen aufgebaut, die über eine Webservice-Schnittstelle kommunizieren, um die Metadaten schrittweise zu verbessern, anzureichern und darzubieten.

Auf der organisatorischen Ebene wird das Projekt die Betreiber von Repositories bei der DINI-Zertifizierung unterstützen und liefert Rückmeldung, wie ihre Güte und Leistungsfähigkeit als Data Provider erhöht werden kann. Auf der technischen Ebene stellt das Projekt einen eigenen Service Provider als Plattform bereit, um neue Mehrwert-Dienste aufzunehmen und die Qualität der bestehenden Dienste zu verbessern. Letztlich dient dies alles dem Ziel, der wissenschaftlichen Gemeinschaft einen Anlaufpunkt zu geben, um zu lesen und gelesen zu werden, wo der in der Open Access Repositories gesicherte Anteil des deutschen Forschungsbeitrags sichtbar wird.

Die größte technische Herausforderung liegt in der Informationsintegration: Metadaten aus heterogenen Quellen

Im System des Service Providers von Open-Access-Netzwerk findet fortlaufend eine solche Wertschöpfungskette in Richtung besserer Metadaten statt. Eine Reihe verteilter, über einen REST-WebService an die zentrale Datenthaltung gekoppelter Dienste bereitet automatisch die durch Harvesting bezogenen Daten auf und reichert sie mit Mehrwertdaten an. Dabei werden unter anderem Klassifikationen hervorgehoben, über eine Analyse ähnliche Objekte zusammengeführt und der zur Volltext-Datei führende Link identifiziert. Die so aufbereitete Datenbasis ist durch eine webbasierte Nutzerschnittstelle für Suche und Browsing, Alerting-Dienste, News-Feeds und OAI-Export zugänglich und bietet verschiedenste Sichten auf den Datenbestand. Um langfristig neue oder verbesserte Dienste zu integrieren, die

Im Projektverlauf hat sich jedoch erneut bestätigt, dass auch mit findigen Algorithmen die integrierten Daten nur bis zu einem gewissen Grad qualitativ hochwertiger sein können als die Originaldaten in den Repositories. Information kann nicht aus nichts entstehen. Eine Beschränkung sind beispielsweise die mit dem angebotenen Metadatenformat abzubildenden Daten. Das ist als größter gemeinsamer Nenner „Dublin Core Simple“. Wie das Format auf Seiten der Repositories konkret angewendet wird, welche Fehler sich bereits bei der Erfassung einschleichen und wie alternative Formate zu integrieren sind – das offenbart viel Handlungsbedarf. Hier muss der technischen Weiterentwicklung eine organisatorische vorausgehen. Die zukünftigen Aufgaben müssen darauf abzielen, das entstehende Netzwerk

zu stärken und die Betreiber der Repositories noch enger in einer aktiven, deutschlandweiten Community zusammenzubringen.

## Fazit

Open Access als Modell wissenschaftlicher Literaturversorgung kann langfristig nur dann die gesetzten Erwartungen erfüllen, wenn dem rechtlich-organisatorischen Rahmen des freien Zugriffs auf elektronische Publikationen auch ein entsprechender, infrastruktureller Überbau zur Seite gestellt wird. Erweiterbare Plattformen mit nutzerorientierten Mehrwert-Diensten und darin eingebundene Open Access Repositories mit möglichst hochwertigen Metadaten fördern durch gezielte Zusammenarbeit wechselseitig ihren Mehrwert und damit den Nutzungsanreiz des Open-Access-Weges, den sie der wissenschaftlichen Community bieten.

## Literatur

- [1] WEBSEITEN DER UB BIELEFELD:  
<http://www.ub.uni-bielefeld.de/biblio/search/help/invisibleweb.htm>;  
Letzter Zugriff 01.02.2009
- [2] DINI-ARBEITSGRUPPE OPEN ARCHIVES INITIATIVE IN DEUTSCHLAND: *Elektronisches Publizieren an Hochschulen, Inhaltliche Gestaltung der OAI-Schnittstelle*. <http://edoc.hu-berlin.de/series/dini-schriften/2-de/PDF/2-de.pdf>; 2003
- [3] LAGOZE, CARL; VAN DE SOMPEL, HERBERT: *The Making of the Open Archives Initiative Protocol for Metadata Harvesting*. <http://public.lanl.gov/herbertv/papers/The%20Maling%20of%20the%20Open%20Archives%20Initiative.pdf>; 2002
- [4] DOBRATZ, SUSANNE; SCHOLZE, FRANK: *Qualitätssicherung durch das DINI-Zertifikat*. Erstveröffentlichung in ZfBB 54 (2007) 4–5, S.194–198; <http://edoc.hu-berlin.de/oa/articles/rem4Ar8VXZoCA/PDF/27LxBlVkdRfoA.pdf>
- [5] Webseiten des Open-Access-Netzwerk-Projektes bei DINI; <http://www.dini.de/projekte/oa-netzwerk/>; Letzter Zugriff 01.02.2009
- [6] SCHOLZE, FRANK; HORSTMANN, WOLFRAM: *Infrastruktur und Netzwerke für Repositories*. [http://www.zbw.eu/ueber\\_uns/projekte/vascoda/ws\\_2007-10-31/03\\_scholze\\_horstmann\\_kiel\\_so.pdf](http://www.zbw.eu/ueber_uns/projekte/vascoda/ws_2007-10-31/03_scholze_horstmann_kiel_so.pdf); 2007
- [7] PROF. SCHIRMBACHER, PETER; SEVERIENS, THOMAS: *Aufbau eines Netzwerkes von Open Access zertifizierten Repositories auf der Basis des DINI-Zertifikats*. DFG-DINI-Workshop „Förderung der wissenschaftlichen Informationslandschaft in Deutschland - Chancen und Strategien beim Aufbau vernetzter Repositories“; [http://www.dini.de/fileadmin/workshops/oa-netzwerk-2008/Severiens\\_Schirmbacher.pdf](http://www.dini.de/fileadmin/workshops/oa-netzwerk-2008/Severiens_Schirmbacher.pdf); 2008

# Ein Zitationsindex elektronischer Dokumente in institutionellen Repositorien

Frank Havemann | [frank.havemann@ibi.hu-berlin.de](mailto:frank.havemann@ibi.hu-berlin.de)

## Qualitätskontrolle von Open-Access-Dokumenten

Hochenergiephysiker waren von jeher daran gewöhnt, ihre Forschungsergebnisse den Kollegen vor der Publikation mitzuteilen, indem sie *preprints* von Zeitschriftenaufsätzen per Post in alle Welt versandten, die so vor Abschluss der langwierigen Begutachtungsprozedur für die Veröffentlichung in einer Zeitschrift ihre Leser erreichten. Es lag dann nahe, diese rasche Kommunikation noch zu beschleunigen, indem man sie über das Internet laufen ließ, was Paul Ginsparg mit dem *arXiv* 1991 realisierte.<sup>1</sup> Aus den *preprints* wurden *eprints*.

Das *arXiv* wurde Modell für alle web-basierten Repositorien, in die Autoren online frei verfügbare elektronische Dokumente einstellen, ohne dass vorher ihre Relevanz und ihre Qualität von Herausgebern oder Gutachtern geprüft werden. Der Zeitvorteil solcher Systeme gegenüber der üblichen Zeitschriftenpublikation ist daran ablesbar, dass viele der *eprints* in anderen *eprints* zitiert werden, bevor sie als Zeitschriftenaufsätze erschienen sind. Eine von mir zusammen mit Studierenden durchgeführte kleine Studie ergab für Artikel zur theoretischen Hochenergiephysik in anderthalb Jahrgängen von *Physical Review D*, dass sie im Schnitt sieben Monate vor dem Erscheinen des Zeitschriftenheftes in das *arXiv* gestellt worden waren und dass drei Viertel von ihnen zum Zeitpunkt des Erscheinens schon mindestens einmal zitiert worden waren.<sup>2</sup>

Sind also Zeitschriften für Hochenergiephysiker und Forscher anderer Fachgebiete, die das *arXiv* nutzen, überflüssig geworden? Keineswegs, sie dienen aber nicht mehr vorrangig der Kommunikation, sondern erhöhen vor allem – je nach ihrem Ansehen – das Ansehen ihrer Autoren in den wissenschaftlichen Gemeinschaften. Forschungskommunikation läuft schneller ohne Journale und benötigt offenbar keine Begutachtung mittels eines aufwendigen Systems von *peer reviewing*; die Leser selber sind die *peers* – jedenfalls in der theoretischen Hochenergiephysik. In anderen Wissenschaftsgebieten ist Eile bei der Kommunikation nicht immer ein so wichtiges Ziel. Es kommt da nicht auf Priorität an, eher auf das Reifen eines Textes, an dem auch die Gutachter noch mitwirken.

Benötigen Forschende und Lehrende Wissen aus Gebieten, auf denen sie nicht so kompetent urteilen können wie im eigenen Fach, werden sie vor allem auf das gesicherte Wissen, das in begutachteten und publizierten Zeitschriftenaufsätzen dokumentiert ist, zurückgreifen. Dabei kann ihnen ein Zitationsindex die im fremden Gebiet am meisten beachteten Beiträge markieren.

## Wozu Zitationsindizes gut sind

Die Zitierungen der *arXiv eprints* zur Hochenergiephysik werden in der online frei verfügbaren SPIRES-HEP-Datenbank bereitgestellt.<sup>3</sup> Mittels eines solchen Zitationsindex können Leser im Zitationsnetzwerk der Aufsätze navigieren – und

*Ausgehend von der Entstehungsgeschichte von Repositorien elektronischer Dokumente wird die Frage der Qualitätskontrolle im Open-Access-Bereich diskutiert sowie der Mehrwert von Zitationsindizes für Nutzer als Leser und Autoren wissenschaftlicher Dokumente. Details eines möglichen Zitationsindex von Open-Access-Dokumenten in Repositorien deutscher wissenschaftlicher Institutionen werden geschildert.*

*Zitationen, elektronische Dokumente, Qualitätskontrolle, Impact Factor*

1 X in *arXiv* bedeutet den griechischen Buchstaben Chi. Vgl. auch: <http://arxiv.org>

2 <http://www.ibi.hu-berlin.de/~fhavem/E-prints.pdf>

3 <http://library.desy.de/spires>

zwar nicht nur rückwärts in die Vergangenheit über die in den Aufsätzen zitierten Quellen, sondern auch vorwärts zu den ein relevantes Dokument zitierenden Aufsätzen. Eine seitliche Navigation innerhalb einer Zeitschicht ist ebenfalls möglich, nämlich über Links zu Aufsätzen, die gleiche Quellen zitieren – man spricht dann von bibliographischer Kopplung – bzw. zu Aufsätzen, die mit dem Startdokument oft zusammen zitiert, d. h. oft koziert werden. Sowohl Koziolation wie bibliographische Kopplung von Aufsätzen sind Zeichen fachlicher Ähnlichkeit – ganz wie die lexikalische Kopplung über gleiche Terme in den Titeln oder Schlüsselwörtern der Aufsätze. Zitationsbasierte Ähnlichkeitsmaße haben gegenüber den termbasierten einen Vorteil: Sie sind unabhängig von der Sprache, in der die Aufsätze verfasst sind.

Hochzitierte Aufsätze, Bücher und andere Quellen haben offenbar die Aufmerksamkeit vieler Autoren erregt und müssen deshalb nicht nur von einigem Belang sein, sondern auch gewisse Qualitätskriterien erfüllen, denn ganz Schlechtes bleibt unbeachtet (fehlerhafte Ergebnisse regen zwar zu Kontroversen, d. h. zum Zitieren an, aber nur dann, wenn aus den Fehlern etwas zu lernen ist). Andererseits werden oft auch gut geschriebene Publikationen interessanter Resultate niveauvoller Forschung nur wenig zitiert. Die Zitationszahl einer Publikation hängt stark von der Zahl der an ihrem Thema interessierten Autoren ab. Es ist deshalb sinnlos, Zitationszahlen über Fachgebietsgrenzen hinaus zu vergleichen, weil ganz unterschiedlich viele Autoren in den verschiedenen Fachgebieten tätig sind. In hochdotierten, heißen Gebieten wird viel publiziert und damit auch viel zitiert. Die mittlere Zitationsrate in einem Gebiet wird aber am Ende nur durch die mittlere Zahl der zitierten Quellen pro Aufsatz bestimmt.

Wenn Nutzer in einer Zitationsdatenbank auf für sie relevante Dokumente stoßen, werden sie als Erstes die höher zitierten näher ansehen, weil sie zu Recht vermuten, dass diese zu den interessanteren gehören (wobei das Alter der zitierten Dokumente beachtet werden muss, weil Zitierung Zeit benötigt).

Mitglieder von Berufungskommissionen, zum Beispiel, könnten nun versucht sein, Kandidaten anhand der Summe der Zitationszahlen der Publikationen, an deren Abfassung diese beteiligt waren, zu vergleichen. Zitationszahlen von Publikationen unterliegen aber – wie vieles andere in Wissenschaft, Wirtschaft und Gesellschaft – dem Matthäusprinzip: Wer hat, dem wird gegeben (Matthäus 25, 29), das von Robert K. Merton 1968 in die Wissenschaftssoziologie eingeführt wurde [1]. Schon oft zitierte Publikationen werden viel wahrscheinlicher noch mehr Zitationen erhalten, als bisher wenig beachtete Werke gleichen Alters. Einer der Begründer der Bibliometrie, der Wissenschaftshistoriker Derek J. de Solla Price hat 1976 das aus der Biologie bekannte Yule-Modell, eine mathematische Formulierung des Matthäusprinzips, auf Zitationen angewendet und damit nicht nur deren schiefe Verteilung erklärt – viele Artikel werden wenig zitiert, nur wenige erhalten viel Beachtung – sondern auch die mathematische Form der Verteilung, die oft gut durch eine fallende Potenzfunktion beschrieben werden kann [2].

Wirken solche positiven Rückkopplungsmechanismen, entstehen in der Regel schiefe Verteilungen. Für ihre Interpretation ist es angemessener, mit Logarithmen von Kennzahlen zu rechnen, als mit den Kennzahlen selber. Bevor wir jedoch auf (für mathematisch nicht ausreichend Gebildete) wenig anschauliche höhere Rechenarten zurückgreifen, folgen wir lieber dem Vorschlag des kalifornischen Physikers Jorge E. Hirsch, der 2005 einen sehr anschaulichen und einfachen Indikator für die Bedeutung des Lebenswerkes von wissenschaftlichen Autoren vorgeschlagen hat, der sich bei theoretischen wie praktischen Bibliometrikern großer Beliebtheit erfreut [3]. Der Hirsch- oder *h*-Index eines Autors hat den Wert *h*, wenn bisher *h* seiner Werke mindestens *h*-mal zitiert worden sind, während alle seine anderen weniger als *h* Zitationen erhalten haben. Die aktuelle Zahl von Zitierungen der meistzitierten Publikation einer Bibliographie beeinflusst ihren *h*-Index überhaupt nicht; auf eine breite Spitze kommt es an. Der *h*-Index von beliebigen Bibliographien wird in der SPIRES-HEP-Datenbank an-

gezeigt.<sup>4</sup> Dieser Service wurde übrigens dort schon eingerichtet, bevor der *arXiv eprint* von Hirsch zum *h*-Index in einer Zeitschrift erschienen war.

## Ein Zitationsindex für Open-Access-Dokumente

Um in einer bibliographischen Datenbank von einem relevanten Dokument zu weiteren geführt zu werden, um die Bedeutsamkeit der Dokumente wie Autoren zu bewerten, kann also ein Zitationsindex der von den erfassten Dokumenten zitierten Quellen von Nutzen sein. Deshalb ist es sinnvoll, auch die geplante Datenbank von elektronischen Dokumenten in Repositorien deutscher Forschungseinrichtungen durch einen Zitationsindex, von den Vorschlagenden *DOARC* genannt, zu komplettieren.<sup>5</sup> Die Wirkung von Autoren auf ihre jeweiligen Fachgemeinschaften kann aber mit einer nationalen Datenbank nur eingeschränkt auf Kollegen an deutschen Institutionen erfasst werden. Für eine vollständige Erfassung muss auf internationale Datenbanken zurückgegriffen werden.<sup>6</sup> Das Gleiche gilt für die Navigation im Netzwerk der Dokumente.

In den kostenpflichtigen Zitationsdatenbanken wie dem *Web of Science* (von Thomson Reuters) oder *SCOPUS* (von Elsevier) werden jedoch bislang hauptsächlich die Zitationsnetzwerke von Zeitschriftenaufsätzen erfasst, Web-Dokumente spielen dort eine geringere Rolle. *Google Scholar* als freie Datenbank von über das Web zugänglichen wissenschaft-

<sup>4</sup> und auch in anderen Zitationsindizes, wie dem *Web of Science*: [http://thomsonreuters.com/products\\_services/scientific/Web\\_of\\_Science](http://thomsonreuters.com/products_services/scientific/Web_of_Science)

<sup>5</sup> Die Datenbank elektronischer Dokumente wird im DFG-Projekt OA Netzwerk installiert (s. den Beitrag von Malitz & Klatt-Kafemann in diesem Heft). *DOARC* (Distributed Open Access Reference Citation Service) ist neben den Projekten OA Netzwerk und OA Statistik ein weiteres von DINI koordiniertes DFG-Projekt zu institutionellen Repositorien (<http://www.dini.de/projekte>).

<sup>6</sup> Internationale Wirkung wird auch an Download-Zahlen sichtbar, die auch schneller verfügbar sind als Zitationszahlen (vgl. auch <http://www.citebase.org>). Um Downloads Evaluationen zugrunde legen zu können, muss Manipulation durch die Autoren ausgeschlossen werden. S. Henneberger schildert in ihrem Beitrag zu diesem Heft den geplanten Dienst zur Statistik von Downloads aus den an der bibliographischen Datenbank beteiligten Repositorien (Projekt OA Statistik).

lichen Dokumenten ermittelt die Metadaten von Publikationen und der in ihnen zitierten Quellen oft noch fehlerhaft. Allerdings stammen viele der Fehler in Quellen-Metadaten von den zitierenden Autoren selber, welche oft keines der verfügbaren Literaturverwaltungssysteme benutzen, mit denen Metadaten effektiv gepflegt werden können.

Exakte und vollständige Metadaten aller in der Zitationsdatenbank DOARC indexierten Publikationen und der in ihnen zitierten Quellen sollen deshalb den Autoren so bereitgestellt werden, dass sie in verschiedenen Formaten in die in Arbeit befindlichen Publikationen eingefügt bzw. in Literaturverwaltungssysteme importiert werden können. Autoren sollen sich außerdem an der Pflege der Metadaten beteiligen können. So tragen sie dazu bei, dass ihre Publikationen international sichtbar werden.

Exakte Metadatensätze sollten für kommerzielle Zitationsdienste wie *Google Scholar* ein willkommener Input sein. Im Gegenzug könnte vom DOARC-Service ein Link zu der *Google-Scholar*-Anzeige der ein *Open-Access*-Dokument zitierenden Aufsätze gesetzt werden, um die oben erwähnte nationale Beschränkung bei Zitationsanalysen zu überwinden.

## Nutzungsoptionen und Datenstruktur

Wie Nutzer im Zitationsnetzwerk navigieren könnten, zeigt ein Demonstrationsmodell des 'Zitationsindex' DOARC.<sup>7</sup> Von einem relevanten Dokument gelangt man nicht nur über die üblichen Links zu zitierten und zitierenden Dokumenten, sondern auch über die Option *Ähnliche Dokumente* zu solchen, die gleiche Quellen zitieren, die also mit dem Startdokument bibliographisch gekoppelt sind. Die bibliographische Kopplung sollte im DOARC-Zitationsindex durch Kozitation und lexikalische Kopplung ergänzt werden. Die ähnlichen Dokumente werden im Modell als Netzwerk-Graph visualisiert, in dem

die Kopplungsstärken an der Liniendicke der Links, die Zitationszahlen der Dokumente an der Größe und ihr Alter an der Färbung der Knoten ablesbar sind. Eine optimale Gestaltung der Navigationsoptionen kann aber nur durch Interaktion mit den Nutzern erreicht werden. Denkbar ist zum Beispiel, dass auch die im parallelen Projekt ermittelten Download-Zahlen im Netzwerk-Graph visualisiert werden.

Stoßen Leser auf für sie interessante Dokumente, dann sollte die Visualisierung der im Zitationsnetzwerk benachbarten und damit ähnlichen Dokumente (mitsamt der Zitations- und Download-Zahlen) für sie von Nutzen sein. Wegen der unterschiedlichen Zitationsgewohnheiten in den Fachgebieten sind auch nur zwischen fachlich benachbarten Artikeln gleichen Alters Zitations- und Download-Zahlen sinnvoll vergleichbar.

Aus den Dokumenten aller Daten liefernden institutionellen Repositorien werden in einem automatischen Verfahren die Metadaten jeder zitierten Quelle extrahiert und entweder mit einem vorhandenen Metadatensatz identifiziert oder als Repräsentation eines neuen Dokuments etabliert. Dadurch vergrößert sich der Dokumentenraum des geplanten Dienstes beträchtlich, wenn auch nur virtuell, weil viele der zitierten Quellen nicht als *Open-Access*-Dokumente in einem der teilnehmenden Repositorien vorhanden sein werden.

## Motivierung der Autoren

Autoren sollen im DOARC-Service die aus ihren Publikationen extrahierten Metadaten und möglicherweise auch Falschzitierungen ihrer Publikationen durch andere Autoren online korrigieren können. Wie kann man sie dazu bringen, das zu tun? Welche Motive haben sie überhaupt, ihre Texte in ein institutionelles Repository einzustellen? Die schnelle, direkte und kostenlose Verbreitung ihrer Forschungsergebnisse in der Fachgemeinschaft kann bei fachlich organisierten Repositorien wie dem *arXiv* ein hinreichender Grund sein, die zwar geringe, aber doch lästige Mühe des Hochladens auf sich zu nehmen, aber nicht bei Re-

positorien, die an Institutionen angebunden sind. Was ist hier der Mehrwert für die Autoren?

Der entworfene Zitationsdienst DOARC wird nach seiner Realisierung Autoren dadurch nützen, dass er Metadaten aller Dokumente und der in ihnen zitierten Quellen für die Zitierung in unterschiedlichen Formaten und als Input für Literaturverwaltungssysteme bereitstellt. Das hilft Autoren beim Schreiben, aber motiviert sie noch nicht, das Geschriebene auch in ein institutionelles Repository einzustellen.

Ein gewichtiges Motiv kann am Ende nur die bessere Sichtbarkeit und Zugänglichkeit der *Open-Access*-Dokumente in institutionellen Repositorien sein, welche ihre Metadaten in die bibliographische Datenbank einspeisen. Dadurch, dass alle Dokumente der beteiligten Repositorien mit ihren zitierten Quellen über eine Oberfläche erreichbar sein werden, werden nicht nur Nutzer als Leser und Autoren angezogen, sondern auch andere Zitationsdienste wie *Google Scholar*, die Metadaten importieren können. Indem andererseits Navigationsmöglichkeiten im Zitationsnetzwerk erprobt werden, die *Google Scholar* und andere Zitationsdatenbanken so noch nicht anbieten, können wir mit der Realisierung dieses Vorhabens dazu beitragen, Zitationsdienste komfortabler zu machen.

## Danksagung

Für wertvolle Hinweise zur Verbesserung des Texts danke ich W. Christen, E. Hilf und P. Schirnbacher.

## Literatur

- [1] MERTON, R.: *The Matthew Effect in Science*. *Science* 159 (3810), 1968, S. 56–63.
- [2] DE SOLLA PRICE, D.: *A General Theory of Bibliometric and Other Cumulative Advantage Processes*. *JASIS* 27, 1976, S. 292–306.
- [3] HIRSCH, J. E.: *An index to quantify an individual's scientific research output*. *PNAS* 102 (46), 2005, S. 16569–16572. <http://arxiv.org/abs/physics/0508025>.

<sup>7</sup> Vgl. <http://doarc.projects.isn-oldenburg.de>. Das Modell wurde von E. Hilf, Th. Severiens, W. Christen, M. Maune und mir für die Begutachtung unseres bei der DFG eingereichten Projekts DOARC entwickelt.



# Impact von Open-Access-Publikationen

Sabine Henneberger | shenneberger@cms.hu-berlin.de

## Der Impact wissenschaftlicher Publikationen

Im Zusammenhang mit wissenschaftlichen Publikationen spricht man von Impact, wenn man die, vor allem kurzfristige, Wirkung in Fachkreisen meint. Diese, so die Annahme, wird widergespiegelt, wenn sich Autor(inn)en auf vorangegangene Publikationen beziehen, indem sie diese zitieren. Um ein Maß für den Impact zu bilden, wird zumeist die Anzahl der Zitationen in nachfolgenden Publikationen hinzugezogen. Ein Grund dafür ist, dass sich Zitationen einfach durch Zählung messen lassen. Die Bedeutung, die auf Zitationen beruhende Impact-Maße als Wissenschaftsindikatoren bei der Beurteilung von Publikationen und damit der Leistungsfähigkeit von Autorinnen und Autoren bis heute erlangt haben, hängt mit dieser einfachen Möglichkeit, die durch elektronische Datenverarbeitung auf Datenbanken gestützt automatisch durchgeführt wird, zusammen, andererseits aber auch mit der Annahme, dass mit Zitationen ein Belohnungssystem in der Wissenschaft realisiert wird [1].

Ausführlicheres über Ideen, Hintergründe und unterschiedliche Ansätze der Zitationsanalyse findet man im vorangegangenen Artikel dieses Journals.

Obwohl auf Zitationen beruhende Impact-Maße (Citation Impact) umstritten und ihre Verwendung als ausschließliches Qualitätsmerkmal äußerst fragwürdig ist, spielen sie doch eine herausragende Rolle bei der Bewertung wissenschaftlicher Leistung. Dem liegt die Idee zugrunde, dass Zahlen, gewonnen aus bibliometrischen Analysen, also Statistiken von

Zitationen, objektiv und deshalb besser geeignet sind als z. B. die Einschätzung von Forschungsleistungen durch Peer Review. Zusätzlich dazu, dass Zahlen nur einen Teilaspekt wissenschaftlicher Leistungen abbilden können, werden Statistiken einerseits häufig falsch angewendet und andererseits Ergebnisse falsch interpretiert. Die Folge davon ist, dass die vermeintliche Objektivität der Zahlen oft nur eine Illusion ist. Zu dieser Aussage gelangen die Autorinnen und Autoren des IMU<sup>1</sup>-Reports, in welchem eine umfassende kritische Auseinandersetzung mit der Zitationsanalyse geführt wird [2].

Eine Alternative ist die Verwendung von Nutzungsdaten von Publikationen zur Bewertung des Impacts. Nutzungsdaten können von traditionellen, also ausschließlich gedruckten Publikationen erhoben werden, wesentlich besser aber von digitalen Publikationen, die hier im Mittelpunkt stehen.

Die Nutzung digitaler Quellen, die durch elektronische Systeme protokolliert wird, kann auf der Basis von Netzwerkmodellen, die aus den Sozialwissenschaften bekannt sind, analysiert werden. Beispiele sind der PageRank-Algorithmus (Google<sup>2</sup>) oder die Analyse von Nutzungsdaten in Form von „Usage Networks“ oder „Reader Generated Networks“ [3,4]. Es gibt eine Reihe von Vorschlägen, wie auf der Basis der Analyse von Nutzungsdaten Maße für den Impact von Publikationen definiert werden können.

Solche Alternativen zum Citation Impact sind die von Bollen et al. vorgeschlagenen Metriken für Journale [5].

*Impact wissenschaftlicher Publikationen bedeutet nicht nur Citation Impact. Auch der Download Impact kann als ein Maß dafür angesehen werden, wie eine Publikation in Fachkreisen Verbreitung findet. Die Bestimmung des Download Impact aus Protokollen elektronischer Systeme, die Open-Access-Publikationen über das Internet anbieten, ist jedoch nicht unproblematisch und muss normiert werden. Das DFG-Projekt „OA-Statistik“ hat sich dieser Aufgabe gestellt.*

*Citation Impact, Download Impact, Statistik, Zitationsanalyse*

1 IMU: International Mathematical Union

2 Google: <http://www.google.de>

Nutzungsdaten sind unter anderem die Downloadzahlen von digitalen Publikationen. Jeder Zugriff eines Internetnutzers auf eine Website, speziell auf eine Publikation mittels eines Browsers, ist solch ein Download, egal, ob die Daten nur im Browser sichtbar sind oder gespeichert werden. Die Begriffe Downloadzahl und Zugriffszahl werden synonym verwendet. Die Anzahl der Downloads einer digitalen Publikation wird als Download Impact bezeichnet.

Wer auf dem edoc-Server der Humboldt-Universität publiziert, kann sich „seine“ Downloadzahlen ansehen. Sie werden auf der sogenannten „Frontdoor“, der Eingangsseite zur Publikation, veröffentlicht. Die „Frontdoor“, auch als „Abstract“ bezeichnet, erreicht man z. B. über die Homepage des edoc-Servers<sup>3</sup> im Bereich „Lesen“ oder über die Suche.

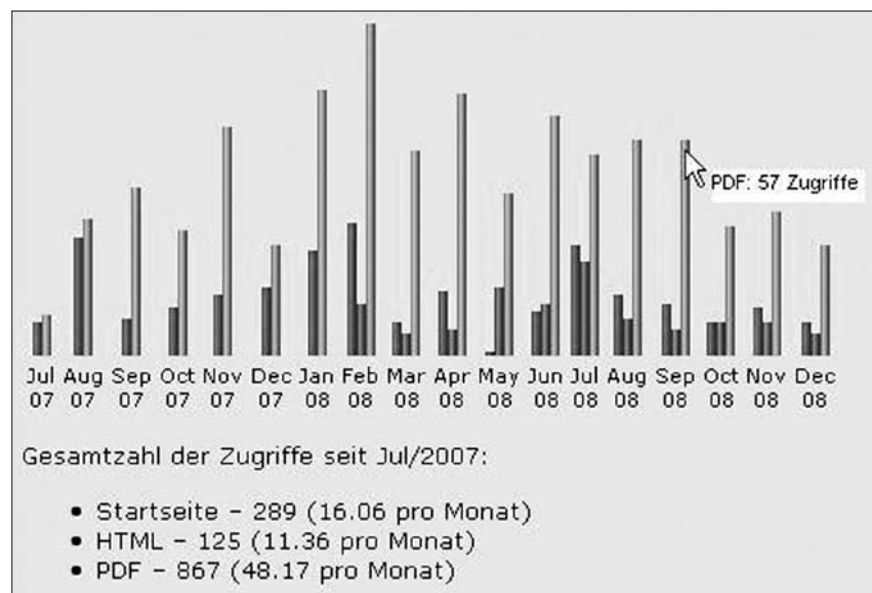


Abb. 1: Zugriffsstatistik auf dem edoc-Server der Humboldt-Universität

Diese Anzeige der Downloadzahlen ist ein Service für die Autorinnen und Autoren, welcher bei weitem nicht von allen Open Access Repositories geboten wird.

Was sagen diese Zahlen aber aus und sind sie wirklich relevant für Autorinnen und Autoren?

## Impact und Open Access

Gibt es einen Zusammenhang zwischen Download Impact und Citation Impact? Diese Frage werden Autorinnen und Autoren, besonders wenn sie auf einem Open Access Repository publizieren, stellen, wird ihre wissenschaftliche Leistung doch nach wie vor auf Grundlage von Zitationen in nachfolgenden Publikationen gemessen.

Brody et al. wiesen die Korrelation von Download Impact (Zugriffshäufigkeit), gemessen ab dem Zeitpunkt der Veröffentlichung auf arXiv.org<sup>4</sup>, einem Open Access Repository für Physik, Mathematik, Informatik, Biologie und Statistik, und späterem Citation Impact (Zitationshäufigkeit) von Artikeln auf einigen Gebieten der Physik und Mathematik nach und folgern daraus, dass der

bald vom „Open Access Advantage“ im Hinblick auf den Citation Impact, zumindest für Artikel, die Rede.

Allerdings kann diese Aussage nicht in der Allgemeinheit, wie sie oft verwendet wird, aufrecht erhalten werden.

So wurde der Einfluss von Open Access auf den Citation Impact von Artikeln eines astrophysikalischen Journals untersucht. Die Autoren stellten fest, dass die Open-Access-Publikation keinen höheren Citation Impact zur Folge hatte. Sie kommen zu dem Schluss, dass es viele Argumente für Open Access gibt, von einem generellen „Open Access Citation Advantage“ aber nicht die Rede sein kann [7].

Ebenso beobachteten Davis et al. zwar wesentlich größere Downloadzahlen für Open-Access-Artikel, nicht aber höhere Zitationszahlen [8].

Diese letzte Arbeit führte zu heftigen Kontroversen<sup>5</sup>, nicht zuletzt in methodischer Hinsicht, die einen Einblick in die oft auch polemisch geführte Diskussion zum Thema gibt.

Eine gute Möglichkeit, die Entwicklung auf diesem Gebiet zu verfolgen, bietet die Website des Open Citation Project.<sup>6</sup>

Es scheint im Moment nicht sicher, ob die Veröffentlichung auf einem Open Access Repository dazu führt, dass eine Publikation häufiger zitiert wird, als würde sie nur Non Open Access veröffentlicht werden. Es gibt Beispiele, die dafür, und solche, die dagegen sprechen.

Aber es ist klar, dass Open Access zu höheren Downloadzahlen und damit zu einer besseren Verbreitung wissenschaftlicher Ergebnisse führt. Impact kann eben nicht nur im Sinne von Citation Impact verstanden werden.

<sup>5</sup> siehe [http://www.bmj.com/cgi/eletters/337/jul31\\_1/a568#199775](http://www.bmj.com/cgi/eletters/337/jul31_1/a568#199775), gelesen 14.8.2008

<sup>6</sup> The effect of open access and downloads ('hits') on citation impact: a bibliography of studies, <http://opcit.eprints.org/oacitation-biblio.html>, gelesen 22.12.08

<sup>3</sup> <http://edoc.hu-berlin.de>

<sup>4</sup> <http://archivx.org/>

## Das Problem des Download Impacts

Was verbirgt sich genau hinter den Downloadzahlen?

Download- oder Zugriffszahlen einer Publikation werden ermittelt, indem Protokolle, die so genannten Logfiles des elektronischen Systems, mit welchem die Publikation im Internet angeboten wird, ausgewertet werden.

Dafür existiert eine Vielzahl von Programmen, die Downloadstatistiken von Repositories durchführen und eine Gemeinsamkeit besitzen:

Die Downloadzahl einer Publikation ergibt sich aus der Anzahl der Datensätze des Logfiles, die einen Zugriff auf diese Publikation repräsentieren und nicht durch Algorithmus oder Konfiguration ausgeschlossen werden.

Diese Algorithmen differieren erheblich, so dass nicht davon ausgegangen werden kann, dass Ergebnisse unterschiedlicher Programme vergleichbar sind. Selbst bei der Verwendung des gleichen Analyseprogramms für verschiedene Repositories sind durch die vielfältigen Konfigurationsmöglichkeiten keine verlässlichen Aussagen darüber zu treffen, ob die Downloadhäufigkeiten auf die Dokumente des einen größer als auf die des anderen Repository sind.

Von besonderer Bedeutung ist die Filterung maschineller Zugriffe. Solche Robot-Zugriffe machen oft mehr als 50% alle Zugriffe aus und sollen natürlich möglichst ausgeschlossen werden. Dafür gibt es eine Reihe von Prinzipien, nach denen vorgegangen wird. Außerdem existieren Listen bekannter Robots, anhand derer ein Robot ausgeschlossen werden kann, wenn er sich zu erkennen gibt. Sowohl die Verwendung von Robotlisten als auch die Befolgung von Prinzipien zur Robotererkennung können keinen vollständigen Ausschluss maschineller Zugriffe garantieren. Es kann immer nur eine möglichst gute Annäherung erreicht werden.

Solche Downloadzahlen, die stark davon abhängen, unter welchen Bedingungen sie ermittelt wurden, taugen kaum als Maß eines Download Impacts.

Ein Versuch, Downloadzahlen vergleichbar zu machen, wird im Projekt

COUNTER<sup>7</sup> unternommen. Dieses Projekt startete im Jahr 2002 mit dem Ziel, den damaligen Hauptakteuren unter den Anbietern von Online-Ressourcen, Herausgebern und Bibliothekaren, verlässliche Zahlen über die Nutzung ihrer Produkte und Bibliotheksbestände zu bieten. Die dort aufgestellten Richtlinien bilden inzwischen einen gewissen Standard, reichen aber für die Normierung der Downloadstatistik für Open Access Repositories nicht aus, da zum Beispiel die Filterung von Robotzugriffen hier keine Rolle spielt. Weitere Normierungsbestrebungen für Downloadstatistiken werden z. B. im Projekt LogEc<sup>8</sup> und durch die IFBAC<sup>9</sup> unternommen.

Der aktuelle Situation für Open Access Repositories ist so, dass die Downloadzahlen von Publikationen innerhalb eines Repository verglichen werden können, da sie unter identischen Bedingungen ermittelt wurden und sich ein Download Impact immer nur auf genau dieses eine Repository bezieht.

Ein Maß für den Download Impact einer Publikation kann so aber nicht bestimmt werden. Es muss als Mindestanforderung möglich sein, Downloadzahlen so zu ermitteln, dass sie als Download Impact einer Gruppe von Repositories angesehen werden können.

Für Autorinnen und Autoren bedeutet dies im Moment, dass die angegebenen Downloadzahlen Repository-abhängig sind. So können auch die durchschnittlichen Downloadzahlen eines anderen Repository höher oder niedriger ausfallen. Sie können aber mit den Downloadzahlen anderer Publikationen des gleichen Repository verglichen und als Hinweis auf die Verbreitung der Publikation dienen.

## Eine Lösung

Downloadstatistiken von Open Access Repositories vergleichbar zu machen, ist das Anliegen des DFG-Projektes "Infra-

struktur für standardisierte Nutzungsstatistiken unter besonderer Berücksichtigung Institutioneller Repositorien"<sup>10</sup>, kurz OA-Statistik. In diesem Projekt arbeiten 4 Institutionen zusammen, die DINI-zertifizierte Repositories<sup>11</sup> mit unterschiedlicher Software betreiben. Das sind DSpace, OPUS und edoc, die Software des edoc-Servers der Humboldt-Universität. Weiterhin kommen Linkresolver, die bibliographische Daten mit Volltexten verknüpfen (SFX<sup>12</sup>), und Lizenzserver zum Einsatz.

Um den international aktuellen Stand bei der Ermittlung von Downloadstatistiken zu diskutieren, wurde an der Humboldt-Universität ein Workshop<sup>13</sup> durchgeführt, an welchem viele namhafte Vertreter von Projekten und Institutionen, die sich mit dem Thema beschäftigen, teilnahmen.

Im Projekt „OA-Statistik“ wird folgendermaßen vorgegangen: Die entstehenden unterschiedlichen Protokolle, die Logfiles, werden lokal, also in der beteiligten Institution, sowohl standardisiert als auch individuell aufbereitet. Standardisiert wird das Verfahren zur Unterscheidung der „Usage Events“, individuell sind die Verfahren zur Zuordnung der „Usage Events“ zu Publikationen, was in den unterschiedlichen Datenstrukturen der Repositories begründet ist. Für die Aufbereitung wird unterschiedliche Software verwendet, im Moment ist der Einsatz von Perl und PHP geplant. Die so aufbereiteten Usage Events werden in Form von OpenURL ContextObjects<sup>14</sup> einem zentralen Aggregator zur Verfügung gestellt. Dieser Aggregator übernimmt nun die eigentliche Arbeit, die Ermittlung der Downloadzahlen für die Publikationen aller 4 beteiligten Institutionen unter gleichen Bedingungen. Hierbei sollen die bereits vorhandenen Standards weitestgehend angewendet werden. Die so bestimmten Download-

<sup>10</sup> <http://www.dini.de/projekte/oa-statistik/>, gelesen 21.01.09

<sup>11</sup> DINI: Deutsche Initiative für Netzwerkinformation e.V., <http://www.dini.de>, gelesen 21.01.09

<sup>12</sup> SFX: <http://www.exlibrisgroup.com/sfx.htm>, gelesen 17.2.2008

<sup>13</sup> <http://www.dini.de/veranstaltungen/workshops/jisc-workshop-2008/>, gelesen 22.01.2009

<sup>14</sup> <http://www.exlibrisgroup.com/category/sfxopenurlsyntax>, gelesen 22.01.2009

<sup>7</sup> COUNTER: Counting Online Usage Of Networked Electronic Resources, <http://www.projectcounter.org>, gelesen 17.2.2008

<sup>8</sup> LogEc: <http://logec.repec.org/>, gelesen 28.12.08

<sup>9</sup> IFABC: International Federation of Audit Bureau of Circulations <http://www.ifabc.org/standards.htm>, gelesen 22.12.08

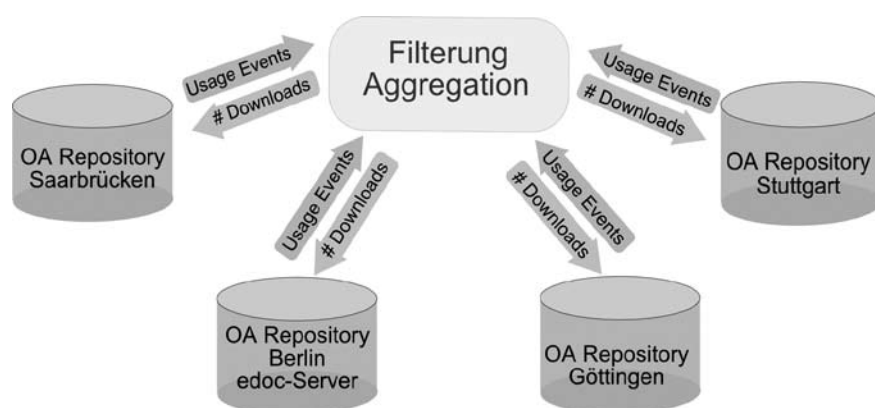


Abb. 2: Überblick über das DFG-Projekt „OA-Statistik“

zahlen der Publikationen werden an die einzelnen Repository-Betreiber zurückgeliefert und in den Metadatenbestand aufgenommen. Von dort aus werden sie über das Internet als zusätzlicher Service für die Autorinnen und Autoren angeboten.

Aber nicht nur die technische Realisierung ist Gegenstand des Projektes. In einem eigenen Arbeitspaket werden Fragen des Datenschutzes geklärt. Das Projekt OA-Statistik kooperiert außerdem eng mit dem DFG-Projekt „Aufbau eines Netzwerks zertifizierter Open Access-Repositories“<sup>15</sup> und liefert mit dem Anbieten von Downloadzahlen einen wichtigen Mehrwertdienst.

Die Ergebnisse des Projektes, mit denen im Frühjahr 2010 zu rechnen sein wird, sollen einerseits in Form von Empfehlungen in das DINI-Zertifikat aufgenommen werden, um eine möglichst weite Verbreitung zu erreichen, andererseits soll die entstandene Software einschließlich Anleitungen zu ihrem Gebrauch Betreibern weiterer Repositories zur Verfügung gestellt werden, die dann ebenfalls in die Lage versetzt werden können, ihren Autorinnen und Autoren vergleichbare Downloadzahlen anzubieten.

## Literatur

- [1] ADLER, ROBERT; EWING, JOHN UND TAYLOR, PETER: *Citation Statistics*. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), 2008, <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>, gelesen am 28.12.08
- [2] BOLLEN, JOHAN; LUCE, RICK; VEMULAPALLI, SOMA SEKHARA UND XU, WEINING: *Usage Analysis for the Identification of Research Trends in Digital Libraries*. D-Lib Magazine Volume 9, Number 5, 2003, [ISSN 1082-9873]; <http://www.dlib.org/dlib/may03/bollen/05bollen.html>, gelesen am 22.01.2009
- [3] BOLLEN, JOHANN UND VAN DE SOMPEL, HERBERT: *Mapping the structure of science through usage*, *Scientometrics* 69, No. 2, 2006, S. 227–258.
- [4] BOLLEN, JOHANN; VAN DE SOMPEL, HERBERT; SMITH, JOAN UND LUCE, RICK: *Toward alternative metrics of journal impact: A comparison of download and citation data*. 2005, <http://arxiv.org/abs/cs.DL/0503007>, gelesen am 4.7.08
- [5] BRODY, T.; HARNAD, S. UND CARR, L.: *Earlier Web Usage Statistics as Predictors of Later Citation Impact*, *Journal of the American Association for Information Science and Technology*. (JASIST) 57(8) 57(8), 2006, S. 1060–1072.
- [6] DAVIS; LEWENSTEIN; SIMON; BOOTH UND CONOLLY: *Open access publishing, article downloads, and citations: randomised controlled trial*. 2008, [http://www.bmj.com/cgi/content/abstract/337/jul31\\_1/a568?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=citation&searchid=1&FIRSTINDEX=0&volume=337&resourcetype=HWCIT](http://www.bmj.com/cgi/content/abstract/337/jul31_1/a568?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=citation&searchid=1&FIRSTINDEX=0&volume=337&resourcetype=HWCIT), gelesen am 13.8.2008
- [7] HORNBOSTEL, STEFAN: *Wissenschaftsindikatoren: Bewertungen in der Wissenschaft*. Westdeutscher Verlag, Opladen, 1997, ISBN: 3-531-12908-2
- [8] KURTZ, MICHAEL J. UND HENNEKEN, EDWIN A.: *Open Access does not increase citations for research articles from The Astrophysical Journal*. 2007, <http://arxiv.org/abs/0709.0896>, gelesen am 22.01.2009

<sup>15</sup> <http://www.dini.de/projekte/oa-netzwerk/>

## Vorbemerkung

Jeder Autor besitzt an dem von ihm produzierten Werk das Urheberrecht. Das Urheberrecht beinhaltet grundsätzlich, dass der Autor einer Veröffentlichung, Bearbeitung, Verwertung und Kopie zustimmen muss [1]. Es erlaubt in begrenztem Umfang Kopien einerseits für den privaten Gebrauch und andererseits im Dienste von Forschung und Bildung [2]. Das Urheberrecht garantiert jedem Autor das Recht an seinem Werk und den Anspruch auf dessen Verwertung für die Dauer seines Lebens und in Deutschland auch noch für 70 Jahre nach seinem Tod. Das Urheberrecht verbleibt immer beim Autor und kann von diesem nicht abgetreten werden. Unberührt davon ist das Verwertungs- bzw. Nutzungsrecht. Der Autor kann die Verwertung seines Werkes an Dritte übertragen, z. B. einen Verlag, und diesem dabei einfache oder exklusive Nutzungsrechte einräumen und dafür eine entsprechende Kompensation vereinbaren. Die im angelsächsischen Raum übliche Copyright-Klausel, die gerade im Internet auch auf deutschsprachigen Webseiten oft zu lesen ist, kann aber so nicht auf den deutschen Rechtsraum angewandt werden. Das Copyright hat nichts mit dem Urheberrecht zu tun, sondern mit dem „right-to-copy“, entspricht also eher dem deutschen Nutzungs- und Verwertungsrecht, das *Copyright* nennt also den Inhaber der Verwertungsrechte. Eine beispielhafte Anwendung des Urheberrechts ist die Nachnutzung eines Bildes aus einer Zeitung. Kopien oder Bearbeitungen dieses Bildes dürfen nur nach entsprechender Genehmigung des Urhebers bzw. des vertraglich dazu er-

mächtigten Verlegers publiziert werden. Zusätzlich muss in der Regel eine Abstimmung bezüglich eines finanziellen Entgelts erfolgen. Da diese Abstimmungen zeitlichen und finanziellen Aufwand darstellen und Verwertungsansprüche durchaus auch strittig sein können, wurde im Jahr 2001 in den USA von Lawrence Lessig und anderen *Creative Commons* gegründet. Das Hauptziel von *Creative Commons* ist die Schaffung verschiedener Lizenzen, die das Urheberrecht insbesondere durch ihre verschiedenen Schreibweisen allgemein verständlicher machen und dem Nutzer der Lizenzen explizit erlauben, gewisse seiner urheberrechtlich garantierten Rechte bewusst freizugeben. Dabei ist darauf hinzuweisen, dass es sich um die bedingte Freigabe allen geistigen Eigentums handeln kann; neben textgebundenen Werken ist dies auch auf andere mediale Daten übertragbar. Durch diese Förderung des Verständnisses wird eine legale Nutzung bereits existierender Inhalte insbesondere im Internet geschützt und gefördert.

Der Gedanke hinter *Creative Commons* ist, verschiedene Lizenzen zu definieren, die eine Möglichkeit zu schaffen, für ein Werk „einige Rechte vorbehalten“ („some rights reserved“) oder „keine Rechte vorbehalten“ („no rights reserved“) im Gegensatz zum im Urheberrecht verankerten „alle Rechte vorbehalten“ („all rights reserved“) festzulegen. Autoren geben bei einer Nutzung von *Creative Commons* nicht das Urheberrecht an ihren Werken ab, sondern sollen durch *Creative Commons* eine Möglichkeit erhalten, explizit darzustellen, auf welche ihrer Rechte sie verzichten. *Creative Commons* stellen somit ein standardisiertes Vertragsangebot über

*Creative Commons stellen ein Lizenzmodell dar, das dem Autor ermöglicht, seine Werke unter Wahrung seines Urheberrechts zur Nachnutzung und eventuellen Bearbeitung freizugeben. Dafür werden leicht verständliche Beschreibungen der Lizenzen und eine maschinenlesbare Darstellung zur Verfügung gestellt.*

*Creative Commons, Urheberrecht, Lizenzierung*



Verwertungsrechte dar. Damit können Autoren wählen, welche Rechte der Nachnutzung an ihren Werken sie anderen einräumen möchten. Andere hingegen möchten zu weiteren Arbeiten beitragen, diese verändern und wiederum veröffentlichen – sei es, ein Bild oder ein Musikstück zu bearbeiten. Diese Art der Kreativität möchte *Creative Commons* unterstützen und durch eine einheitliche Kennzeichnung und Lizenzierung, insbesondere aber auch durch einfache Suchmöglichkeiten, fördern. Zur Kennzeichnung werden Metadaten, also Informationen über das Werk wie z. B. der Name des Autors, genutzt. Dieser Aspekt wird im nächsten Abschnitt genauer dargestellt.

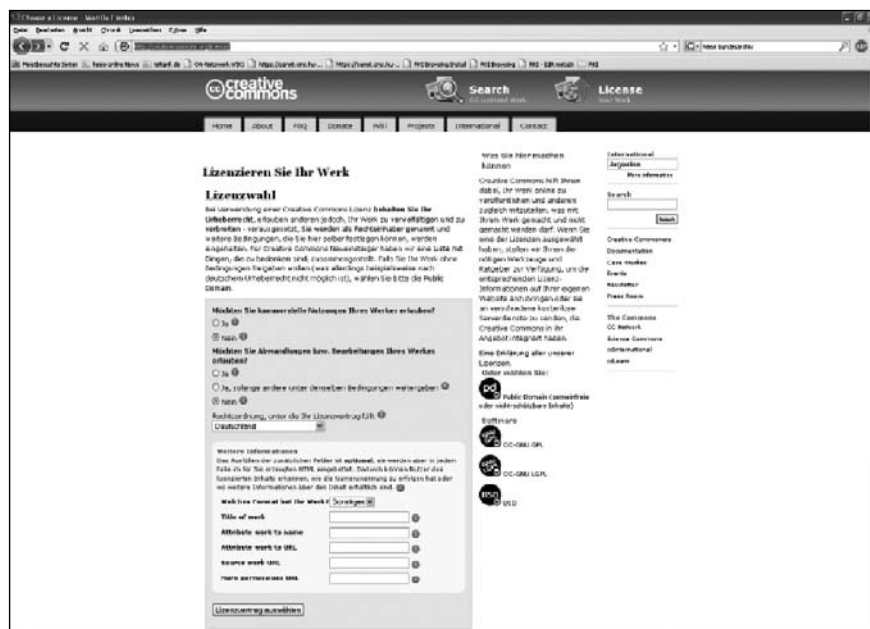


Abb. 1: Erstellung der Creative-Commons-Lizenz

## Auszeichnung

*Creative Commons* hat mehrere Lizenzen entwickelt, die in verschiedenen Landesversionen existieren. Zu jeder Lizenz existieren jeweils drei verschiedene Darstellungen: in einer maschinenlesbaren („digital code“), in einer nicht-juristischen („commons deed“) und in einer juristischen Schreibweise („legal code“). Die juristische Schreibweise umfasst den vollständigen Lizenzvertragstext. Diese Schreibweise stellt das juristische Fundament dar und ist an das Urheberrecht angepasst. Der Vertragstext muss

an das nationale Recht angepasst werden und ist, so weit es rechtlich möglich ist, weltweit inhaltlich gleich. Momentan existieren entsprechende Anpassungen für über 50 Staaten. Federführend in Deutschland ist die *Europäische EDV-Akademie des Rechts gGmbH* [3], die vom Institut für Rechtsinformatik der Universität des Saarlandes unterstützt wird. Die juristische Schreibweise muss gegebenenfalls der sich ändernden Rechtsgrundlage angepasst werden. Daraus resultieren unter anderem auch die Versionsnummern der *Creative-Commons-Lizenzen*. Aktuell ist die Version 3.0, die auch an das kürzlich novellierte deutsche Urheberrechtsgesetz angepasst wurde.

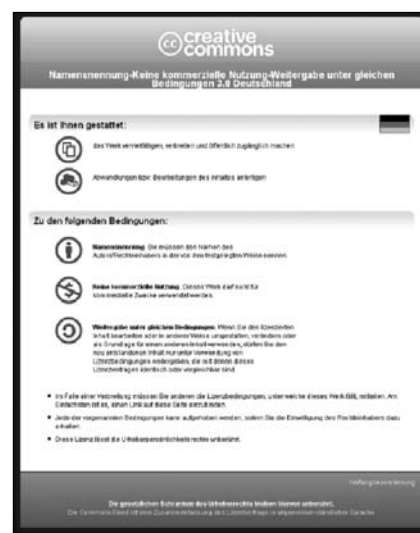


Abb. 2: Erzeugte Lizenz in der nicht-juristischen Schreibweise.

Die maschinenlesbare Schreibweise ist dabei eine Metadaten-Beschreibung in einem XML-Dialekt, die entweder in Form von Resource Description Framework (RDF), als XMP (Extensible Metadata Platform) oder schlicht als HTML-Link auftreten kann. Diese Schreibweise ist besonders wichtig für die Nutzung im Internet. Anhand der Metadaten, die den Titel, den Namen des Autors und die Nutzungsbedingungen des Werkes enthalten, können Suchmaschinen automatisch die Inhalte indexieren und Suchenden Werke mit den gewünschten Urheberrechts-Freigaben präsentieren. Diese Funktionalität ist ein zusätzlicher Nutzen und wird erstmals durch *Creative Commons* ermöglicht. Sie erlaubt es einem potenziellen Nachnutzer, sehr einfach alle frei verfügbaren, bearbeitbaren Objekte zu durchsuchen. Zur korrekten Indexierung müssen die RDF-Daten zusammen mit dem Dokument gespeichert bzw. auf der Webseite integriert werden. Ein Beispiel der RDF-Daten ist in Abbildung 3 zu sehen. Alternativ können diese Metadaten auch im XMP-Format (Extensible Metadata Platform) gespeichert sein. Dieses von Adobe begründete Format eignet sich sehr gut zur Kennzeichnung von PDF-Dateien, da die XMP-Daten direkt im Adobe Acrobat eingebunden und in die PDF-Dateien integriert werden und somit auch leicht von gängigen Suchmaschinen ausgewertet werden können. Weitergehende Informationen werden unter [4] angeführt.

```

<rdf:RDF xmlns="http://web.resource.org/cc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <Work rdf:about="">
    <license rdf:resource="http://creativecommons.org/licenses/by-nc/2.0/de/" />
  </Work>

  <License rdf:about="http://creativecommons.org/licenses/by-nc/2.0/de/">
    <permits rdf:resource="http://web.resource.org/cc/Reproduction" />
    <permits rdf:resource="http://web.resource.org/cc/Distribution" />
    <requires rdf:resource="http://web.resource.org/cc/Notice" />
    <requires rdf:resource="http://web.resource.org/cc/Attribution" />
    <prohibits rdf:resource="http://web.resource.org/cc/CommercialUse" />
    <permits rdf:resource="http://web.resource.org/cc/DerivativeWorks" />
  </License>

</rdf:RDF>

```

Abb. 3: Meta-Daten der Creative-Commons-Lizenz Meta-Daten der Creative-Commons-Lizenz

## Ausprägungen

Prinzipbedingt gibt es nicht nur eine *Creative-Commons*-Lizenz. Je nachdem, welche Rechte vom Urheber freigegeben werden, sind unterschiedliche Lizenzen notwendig. Mögliche „Freigaben“ sind die kommerzielle Verwertbarkeit des Werkes, die Auszeichnung, ob Bearbeitungen erlaubt sind und unter welcher Lizenz das neu entstandene Werk veröffentlicht werden muss.

Unabhängig von den gewählten freigegeben Rechten gelten immer die nachfolgenden Grundrechte: Der Urheber behält in jeder Lizenz das Copyright. Jeder Nachnutzer, der dann als Lizenznehmer auftritt, muss einen Verweis auf die *Creative-Commons*-Lizenz und einen Verweis auf das ursprüngliche Werk angeben. Dafür darf der Lizenznehmer das Werk beliebig kopieren, verbreiten und aufführen. Eine *Creative-Commons*-Lizenz ist für ein Werk nicht mehr widerrufbar. Ein einmal unter einer Lizenz veröffentlichtes Werk ist damit dauerhaft veröffentlicht. Es kann alternativ nur unter einer anderen *Creative-Commons*-Lizenz veröffentlicht werden – dies erscheint jedoch nicht sehr sinnvoll.

### „Nicht kommerziell“

„Nicht kommerziell“ besagt, dass Lizenznehmer des mit diesem Attribut versehenen Werkes die Arbeit zwar kopieren und bearbeiten dürfen, aber eine kommerzielle Nutzung des Werkes ohne vorherige Zustimmung des Urhebers nicht erlaubt ist. Es schließt also die kommerzielle Nutzung nicht per se aus,

sondern verlangt eine explizite Erlaubnis des Urhebers. Der Urheber selbst kann sein Werk also kommerziell vermarkten und es gleichzeitig unter einer *Creative-Commons*-Lizenz veröffentlichen.

### „Keine Bearbeitung“

„Keine Bearbeitung“ beschränkt die Möglichkeiten für Lizenznehmer dahingehend, dass sie das ursprüngliche Werk zwar nutzen und kopieren dürfen, es jedoch in keiner Art und Weise bearbeiten dürfen. Dieses Attribut dient damit primär der Verbreitung des ursprünglichen Werkes.

### „Weitergabe unter gleichen Bedingungen“

Dieses Attribut der *Creative-Commons*-Lizenz verlangt, dass der Lizenznehmer seine unter Nutzung des ursprünglichen Werkes gewonnenen neuen Werke unter derselben Lizenz veröffentlicht wie das ursprüngliche Werk. Damit kann der Urheber eines Werkes sicherstellen, dass seine Schöpfung auch in Bearbeitungen nur unter den von ihm gedachten Rechten genutzt und veröffentlicht werden darf. Diese Ausprägung schließt gleichzeitig die Grundrechte und die Möglichkeit der Bearbeitung ein.

Die verschiedenen Attribute können auch miteinander gekoppelt werden. Einzig „Weitergabe unter gleichen Bedingungen“ und „keine Bearbeitung“ können nicht kombiniert werden, da sie sich inhaltlich ausschließen.

Die Erstellung der entsprechenden Auszeichnungen und die Wahl der Ausprägungen können bequem im Baukas-

tenprinzip über die *Creative-Commons*-Webseite [5] vorgenommen werden. Abbildung 1 veranschaulicht diesen Prozess. Eine in diesem Beispiel erzeugte Lizenz würde in der nicht-juristischen Schreibweise analog Abbildung 2 aussehen.

## Verwendung

*Creative-Commons*-Lizenzen werden von den meisten Urhebern leider recht zögerlich genutzt. Dies liegt häufig an der Unkenntnis der Existenz dieser Lizenzen, zum Teil an der Vertragsgestaltung der Verlage sowie einer geringen Verbreitung von *Creative Commons* in Institutionellen Repositorien. Der edoc-Server der Humboldt-Universität wurde aus diesem Grund kürzlich erweitert, um den Autoren auch die Nutzung von *Creative-Commons*-Lizenzen zu erleichtern (Näheres dazu im nächsten Abschnitt).

Wie bereits zu Beginn des Artikels gesagt, ersetzen *Creative Commons* nicht das Urheberrecht, sondern sind ein Vertragsangebot, in dem die möglichen Nutzungen des Objekts, z. B. des wissenschaftlichen Werks, standardisiert geregelt sind. Die Urheberrechte sind davon unberührt und verbleiben immer beim Autor. Eine Kennzeichnung eines Werkes mit einer *Creative-Commons*-Lizenz ermöglicht jedoch durch die weltweit einheitliche Darstellung quasi eine Aufhebung des skizzierten Unterschieds zwischen dem deutschen Urheberrecht und dem angelsächsischen *Copyright*.

Insgesamt betrachtet hat sich die Nutzung der Lizenzen weltweit in den letzten Jahren jedoch stark vergrößert, was unter anderem der sich veränderten Publikationskultur im Internet und einer Verbreitung des Web-2.0-Gedankens, sowie entsprechenden Weiterbildungsangeboten geschuldet ist. In Österreich gehören die *Creative-Commons*-Lizenzen schon seit 2002 zum Lehrplan einiger Fachhochschulen. Des Weiteren werden Schulmaterialien, die Lehrer und Lehrerinnen am Tiroler Bildungsserver veröffentlichen, unter einer *Creative-Commons*-Lizenz veröffentlicht. Die Verbreitung im Internet lässt sich z. B. mit den folgenden Zahlen belegen: es gibt z. B. weltweit allein auf Flickr über 1,2 Millionen Blu-

menbilder, die unter einer Creative-Commons-Lizenz veröffentlicht sind, allein über 60000 Seiten zum Thema „Open Access“ etc.

Einen wichtigen Schub hat die Creative-Commons-Bewegung durch die Wahlkampagne von Barack Obama erhalten. Dessen Webseite [10] sowie die meisten in Pressemappen zur Verfügung gestellten Bilder wurden mit einer CC-Lizenz versehen, womit die Berichterstattung von Wahlkampfauftritten und über das Wahlprogramm entsprechend leicht von den vielen ihn unterstützenden Web-2.0-Nutzern nachgenutzt werden konnten und das Marketing für ihn erleichtert wurde. Ein Beispiel, dass Creative Commons nicht nur auf Dokumente oder Webseiten beschränkt ist, ist das *Creative Archive* der BBC [7], in dem vielfältige audio-visuelle Inhalte bereitgestellt werden. Dass Werke, die unter einer *Creative Commons*-Lizenz stehen, kommerziell nicht unerfolgreich sein müssen, zeigt die Veröffentlichung des letzten NIN-Albums unter einer CC-Lizenz. Dieses führt bei Amazon die Liste der Verkäufe von Online-Musik an [11].

In Deutschland wird *Creative Commons* derzeit z. B. vom Dokumentenserver der Universität Tübingen [6] genutzt. Dort wird den Autoren die Nutzung von zwei *Creative-Commons*-Lizenzen angeboten. Auch innerhalb der Max-Planck-Gesellschaft oder in den Helmholtz-Gesellschaften [12] wird *Creative Commons* genutzt. Zwei wichtige Entwicklungen des letzten Jahres in dieser Hinsicht sind die Veröffentlichung von 100000 Bildern des Bundesarchivs unter einer Creative-Commons-Lizenz [13] und das Open-Choice-Programm des Springer-Verlags. Dort kann ein Autor gegen Zahlung einer einmaligen Gebühr sein Werk kostenfrei unter einer entsprechenden Lizenz open access anbieten.

Parallel zu den *Creative-Commons*-Lizenzen wurden in Deutschland die sogenannten DiPP-Lizenzen [15] primär für Textwerke wie e-Journals entwickelt und haben vornehmlich in Deutschland Anwendung gefunden. Unterschiede liegen im Detail, insbesondere können differenzierte Rechte für Teile eines Werkes unter einer DiPP-Lizenz vergeben werden, z. B. kann gesagt werden, dass

nur bestimmte Teile verändert werden dürfen. Dies ist mit *Creative Commons* nicht möglich (weitere Informationen hierzu unter [16]).

Zur weiteren Verbreitung kann jeder Autor selbst beitragen, indem er seine Werke unter eine *Creative-Commons*-Lizenz stellt. Eine einfache Auszeichnung mit einer Grafik wie in Abbildung 5 zum Beispiel auf eigenen Präsentationsfolien, Texten etc. macht dies möglich und ist auch in gedruckter Form sichtbar. Alternativ kann natürlich auch explizit auf die Verwendung einer Lizenz hingewiesen werden, z. B. „dieses Dokument ist unter der *Creative-Commons*-Namensnennungs-Lizenz veröffentlicht“.



Abb. 4: ein einfaches Piktogramm, das die Verwendung der *Creative-Commons*-Namensnennungs-Lizenz anzeigt

## Publizieren auf dem edoc-Server

Auf dem edoc-Server der Humboldt-Universität werden vielfältigste Dokumente open access veröffentlicht. Bislang konnte jeder Autor sein in Form einer PDF-Datei veröffentlichtes Werk mit einer *Creative-Commons*-Lizenz kennzeichnen. Da diese Kennzeichnung aber nicht in den Metadaten des edoc-Servers automatisch erfasst werden konnte, war eine entsprechende Ausgabe auf den Webseiten des edoc-Servers bislang nicht möglich. Zur Vereinfachung für die Autoren und insbesondere auch Erhöhung der Sichtbarkeit der so lizenzierten Werke wurden die Eingabemasken entsprechend angepasst. Ein Autor kann jetzt zwischen den Lizenzen *Creative Commons Namensnennung 3.0 Deutschland*, *Creative Commons Namensnennung-Keine kommerzielle Nutzung 3.0 Deutschland* oder *Creative Commons Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Deutschland* wählen. Wählt er einer dieser Lizenzen aus, wird zukünftig auf der Startseite zu seinem Dokument eine entsprechende Information eingeblendet (siehe Abbildung 5), es ist somit für Leser schneller erkennbar. Des Weiteren wird diese Information automatisch in entsprechende

Kataloge und Suchmaschinen übermittelt. Eine zusätzliche Kennzeichnung innerhalb des Dokuments ist dennoch sehr hilfreich, da diverse Suchmaschinen nur diese Informationen auswerten und insbesondere diese Kennzeichnung von den Betreuern des edoc-Servers nicht nachträglich hinzugefügt werden kann, ohne das Dokument zu verändern und damit dessen Integrität zu gefährden. Eine ausführliche Anleitung ist unter [14] zu finden.



Abb. 5: Beispiel für den edoc-Server

## Kritikpunkte

An den *Creative-Commons*-Lizenzen wird, ähnlich wie bei andern *Open-Source*-Lizenzen, deren fehlende Verträglichkeit – sowohl untereinander, als auch in Zusammenhang mit anderen *CopyLeft*-Lizenzen wie den *GNU*-Lizenzen – kritisiert. Die Ausprägung »Weitergabe unter gleichen Bedingungen« schließt zum Beispiel die Verwendung von Werken mit unterschiedlichen *Creative-Commons*-Lizenzen, die alle mit diesem Attribut versehen sind, in einem Werk aus. Ist eines der dabei verwendeten Werke mit dem Attribut „nicht kommerziell“ versehen und ein anderes nicht, so entstehen rechtliche Probleme, da die daraus gebildeten Lizenzen gegensätzlich sind. Ähnlich ist die Problematik im Verhältnis zu anderen *CopyLeft*-Lizenzen. Diese beinhalten zum Teil ebenfalls Merkmale analog zu „Weitergabe unter gleichen Bedingungen“. Diese können wiederum in Gegensatz zu einer *Creative-Commons*-Lizenz stehen, sodass die Verwendung mehrerer Werke anderer Urheber in einem eigenen Werk lizenzrechtlich ausgeschlossen sein kann.

Dieser Kritikpunkt gilt jedoch bei jeglicher spezieller Lizenzierung. Als Lösung bietet sich gegebenenfalls ein erneutes Publizieren des Werkes unter einer anderen Lizenz an, womit die Probleme umgangen werden können, was allerdings nur durch den Urheber selbst erfolgen kann.

## Fazit

*Creative Commons* sind ein viel versprechender Ansatz zur Lösung der zunehmenden Probleme der Nutzung fremder Ressourcen. Sie bieten eine einfache und effektive Möglichkeit für Autoren, ihre Arbeiten elektronisch anzubieten und juristisch sichere Lizenzmodelle zu nutzen. Die Piktogramme dienen der schnellen Veranschaulichung und die elektronische Auszeichnung der Inhalte in Form von Metadaten erleichtert Programmieren von Suchmaschinen und damit auch deren Nutzern das Auffinden von benötigten und wiederverwendbaren Inhalten. In allen großen Suchmaschinen kann explizit nach „freien Inhalten“ gesucht werden, so z. B. bei *Yahoo* [8] oder *Google*. Damit ist es erstmals möglich, nach allen entsprechend ausgezeichneten, frei verfügbaren, nachnutzbaren Werken im Internet zu suchen. Diese Funktion stellt ein Gegenstück zu den *Digital-Rights-Management*-Systemen und geschlossenen Verlagsangeboten dar, die die Verbreitung von Inhalten kontrollieren und deren Bearbeitung und Weitergabe in der Regel untersagen. Somit wird auch die *Open-Access*-Bewegung unterstützt.

## Literatur

- [1] <http://bundesrecht.juris.de/bundesrecht/urhg/>
- [2] <http://www.lehrer-online.de/dyn/9.asp?url=404559.htm>
- [3] <http://www.eear.eu/>
- [4] <http://creativecommons.org/technology/xmp-help>
- [5] <http://creativecommons.org>
- [6] <http://w210.ub.uni-tuebingen.de/dbt/uni/licenses.php>
- [7] <http://creativearchive.bbc.co.uk/>
- [8] <http://search.yahoo.com/cc>
- [9] <http://flickr.com>
- [10] <http://www.change.gov>
- [11] <http://de.creativecommons.org/nin-meistverkaufte-online-album-2008/>
- [12] <http://oa.helmholtz.de/>
- [13] [http://open-access.net/de/austausch/news/news/period/1228611378//anzeige/kooperation\\_mit\\_wikipedia/](http://open-access.net/de/austausch/news/news/period/1228611378//anzeige/kooperation_mit_wikipedia/)
- [14] [http://edoc.hu-berlin.de/e\\_autoren/anleitungcc.pdf](http://edoc.hu-berlin.de/e_autoren/anleitungcc.pdf)
- [15] <http://www.dipp.nrw.de/lizenzen>
- [16] <http://www.libreas.eu/formal/VergleichDPPLCC.pdf>

# Der edoc-Hostingservice für andere Dokumentenserver

Susanne Dobratz | [dobratz@cms.hu-berlin.de](mailto:dobratz@cms.hu-berlin.de)

## Einleitung

Der Betrieb eines eigenen Dokumentenservers bzw. ein Institutional Repositories scheint heutzutage zum „guten Ton“ zu gehören. Aber geht der Trend wirklich dahin, dass jede Institution auch die technischen und organisatorischen, de facto die gesamten finanziellen, Anstrengungen unternehmen muss, um einen eigenen Service inkl. Server aufzubauen? Laut DINI-Website existieren inzwischen ca. 138 Dokumentenserver in Deutschland. Ein Teil dieser existiert nur virtuell als eigenes Portal zu den Publikationen der jeweiligen Organisation. Was ist demnach der Sinn eines Dokumentenservers, oder anders gefragt: Welche Funktion übernimmt ein Dokumentenserver heute innerhalb einer wissenschaftlichen Einrichtung oder eines Verlages? Ein Dokumentenserver spiegelt zum einen durch die Anzahl und Vielfalt der Publikationen die wissenschaftliche Breite und Tiefe einer Organisation wider. Zum anderen spiegelt er ein spezielles Gesicht seiner Einrichtung wider, welches nicht nur die Qualität der wissenschaftlichen Arbeit nach außen darstellt, sondern auch die moderne Nutzung von Informationstechnologien repräsentiert.

den sind dabei vor allem kleinere bis mittlere Forschungseinrichtungen im Raum Berlin-Brandenburg, für die die Bereitstellung eines Institutional Repositories unter der eigenen Domain der Einrichtung und mit Anpassung an das Corporate Design der gehosteten Einrichtung übernommen wird. Die Dokumentenserver erhalten einen eigenen URN-Bereich, siehe [2], sowie einen technischen und bibliothekarischen Fullservice, je nach Wunsch und Angebot. Ausgerichtet ist dieses Angebot so, dass damit das DINI-Zertifikat für Dokumenten- und Publikationsservices 2007 [3] erfüllt wird und gemeinsam mit der Auftraggebenden Institution beantragt werden kann. Die Basis für ein Hosting bildet die Technologie des edoc-Servers. Dies ist eine Eigenentwicklung, die seit 1998 von der Arbeitsgruppe Elektronisches Publizieren von CMS und UB gepflegt und weiterentwickelt wird. An einigen Stellen wird auf existierende, von anderen Anbietern bereitgestellte Produkte zurückgegriffen, so z. B. bei der Realisierung eines Peer-Review-Prozesses. Dahinter steht dann technologisch das Open-Journal-System des Public-Knowledge-Projektes [4]. Folgende Dienstleistungskategorien können dabei abgedeckt werden:

### A) Bereitstellung eines Basispaketes für Dokumentenserver mit folgenden Kernkomponenten

- Edoc-Server-Basistechnologie und Basis-Metadatendatenbank
- Standard-Weboberfläche
- Konfigurationsschnittstelle (Metadateneingabe, Upload) für Herausgeber

### Die Bereitstellung von Dokumentenservern basierend auf der Technologie des edoc-Servers

Auf Basis der edoc-Servertechnologie bietet die Humboldt-Universität in Zusammenarbeit mit der Humboldt-Innovation GmbH [1] die Bereitstellung kompletter Dokumentenserver an. Die Kun-

*Die Humboldt-Universität bietet auf der Basis des edoc-Servers für wissenschaftliche Einrichtungen auch außerhalb der Humboldt-Universität ein technisches Hosting für E-Journals und komplette Dokumentenserver an. Für die betreuten Hostings ist darin eingeschlossen die Umsetzung der im DINI-Zertifikat für Dokumenten- und Publikationsservices 2007 genannten technischen Bedingungen.*

*Dokumentenserver-Hosting  
Institutional Repository*



- Navigation nach Autorennamen und Dokumententiteln
- Navigation nach DDC Subject Headings
- Suche in den Metadaten
- einfache Volltextsuche in den Dokumenten
- OAI-Schnittstelle mit OAI-DC, XMeta-Diss und MARC21-XML
- Webnutzungsstatistik
- dokumentbezogene Nutzungsstatistik
- Einbindung Proprint-Print-On-Demand-System

## B) Bereitstellung zusätzlicher Ergänzungen

Dazu gehören:

- Open Access Upload für Pre- und Postprints
- Creative Commons-Unterstützung
- Einbindung von Social Tagging Systemen
- Import- bzw. Export-Schnittstellen zu weiteren Systemen
- Einbindung von Peer-Review-Systemen bzw. Konferenzsystemen
- Bereitstellung von Dokumentvorlagen und Publikationstechnologien
- Einbindung multimedialer Objekte, wie Podcasts u. Ä.

## Technologische Komponenten

### Bereitstellung einer Weboberfläche für den Dokumentenserver

Bei der Realisierung eines kompletten Dokumentenservers kann das Corporate Design innerhalb gewisser Parameter übernommen und individuell angepasst werden. Es wird die Verwaltung und Darstellung unterschiedlicher Reihen und Publikationstypen, angefangen bei Monografien, über Institutsreihen oder elektronischen Zeitschriften, bis hin zu digitalisierten Reihen, unterstützt.

Die Navigation innerhalb eines Publikationstyps erfolgt über die Autorennamen, die Titel der Zeitschriften bzw. Dokumente und über die vergebenen Schlagworte bzw. Klassifikationskategorien. Standardmäßig werden die DDC Subject Headings (DDCSH) und die Regensburger Verbundklassifikation (RVK) unterstützt. Weitere Navigationsmög-



Abb. 1: Bereitstellung eines kompletten Dokumentenservers am Beispiel des edoc-Servers des Robert-Koch-Instituts

lichkeiten können bei Bedarf realisiert werden. Eine Suche wird zum einen in den Metadaten durchgeführt und zum anderen als Volltextsuche bereitgestellt.

Für jede dieser Publikationsformen kann eine eigene Uploadschnittstelle und ein gesicherter Eingabebereich bereitgestellt werden. Dies ermöglicht es, unterschiedlichen Herausgebern direkten Zugriff und die Verantwortung für ihr Publikationsvorhaben zu geben.

Eine besondere Funktion fällt hierbei der Eingabemaske für den „grünen Open-Access-Bereich“ zu, die ebenfalls bereitgestellt werden kann. Damit ist es Autoren möglich, selbst die eigenen Publikationen, Pre- oder Postprints, auf den jeweiligen Serverbereich zu bringen. Der Serverver-

walter wird nach erfolgreichem Autorenupload über E-Mail informiert und muss diese Publikationen dann noch auf dem edoc-Server freischalten, damit sie auf dem edoc-Server sichtbar wird.

Erfahrungsgemäß verlangt jeder Publikationstyp bzw. jede Reihe die Vergabe sowohl von allgemeinen als auch spezifischen Metadaten. Dabei variiert die Anzahl und Art der zu vergebenden Metadaten mitunter sehr stark. So wird für jedes Publikationsvorhaben gemeinsam mit den Herausgebern individuell auf der Basis eines Kernmetadatensatzes ein geeignetes Modell erarbeitet.



Abb. 2: Beispiel für eine Metadatenseite für ein Dokument

## Schnittstellen zu anderen Systemen und Diensten

Die Technologie des edoc-Servers gestattet es, unterschiedliche Konfigurationen für Metadatenschemata anzu-legen und somit Publikationstyp spezifische Metadaten zu verwalten. So werden z. B. in einer kunstgeschichtlichen elektronischen Zeitschrift andere Klassifikationsinformationen verwendet als in einer Institutsreihe der Mathematik.

Der edoc-Server stellt ebenfalls Schnittstellen zu anderen Systemen bereit. Damit können Daten mit Bibliothekskatalogen (z. B. den Systemen Aleph oder Pica) ausgetauscht werden.

Die OAI-Schnittstelle unterstützt

zum heutigen Zeitpunkt neben dem standardmäßig verlangten Dublin-Core Simple-Format auch das XMetaDiss-Format zur Abgabe digitaler Hochschulschriften und das Xepicur-Format zur Meldung von URNs an die Deutsche Nationalbibliothek. Zudem wird das Proprint Metadatenschema sowie das eprints Application Profile [8] unterstützt.

Die Dokumente auf einem edoc-Server Hosting können bei Bedarf an Social Tagging Systeme angebunden werden, z. B. Connotea (<http://www.connotea.org/>), RawSugar (<http://www.rawsugar.com/>), FURL (<http://www.furl.net/>), Delicious (<http://delicious.com/>). Ein Export der bibliografischen Daten im Endnote oder BibTex-Format ist bereits vorkonfiguriert.

Damit können die vom edoc-Server zitierten Dokumente relativ einfach und strukturiert in eigene Literaturverwaltungsprogramme aufgenommen werden.

Für den gesamten gehosteten Serverbereich stehen sowohl eine Webzugriffsstatistik als auch dokumentenbezogene Zugriffsstatistiken zur Verfügung.

## Bereitstellung von Publikationstechnologien

Den Kernpunkt der edoc-Services bildet die Unterstützung für Autoren und Herausgeber für den Produktionsprozess der Dokumente. So werden Dokumentvorlagen für unterschiedliche Textverarbeitungssysteme, wie Microsoft Office, Staroffice/OpenOffice, LaTeX, Adobe Indesign, FrameMaker oder QuarkXpress speziell für die jeweiligen Publikationsvorhaben angepasst und bereitgestellt. Dies beinhaltet ebenso die Konfiguration und Programmierung von Konvertierungstechnologien nach XML und PDF, um archivierbare Urformen der Dokumente zu erhalten. Ebenso möglich ist die Bereitstellung und Durchführung des Herausgeber- und Autorensupports für interessierte edoc-Hosting-Nutzer.

## Technische Unterstützung von Peer-Review-Verfahren

Für die Realisierung von e-Journals auf dem edoc-Server und zur Unterstützung von Einreichungs- und Begutachtungsworkflows steht auf dem edoc-Server das Open Journal System (OJS) oder zur Unterstützung von Konferenzen das Open Conference Systems (OCS) zur Verfügung.

## On-Demand-Digitalisierung älterer Bestände

Die Digitalisierung älterer Jahrgänge oder Bestände zur Aufnahme auf den edoc-Server kann über den kostenpflichtigen Ebooks-On-Demand-Service (EoD) der Bibliothek der Humboldt-Universität realisiert werden[6]. Zur Visualisierung auf dem edoc-Server werden sowohl PDF-Dateien zum individuellen Ausdruck als auch virtuell blätterbare Bücher aus den Digitalisaten erstellt und veröffentlicht.

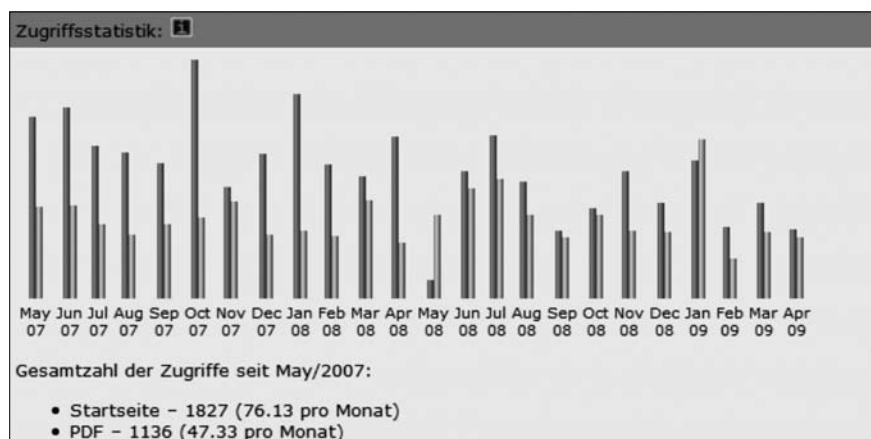


Abb. 3: Beispiel für eine dokumentenbezogene Zugriffsstatistik für einen Artikel des Publikationsvorhabens

ATHENÄUM

HOME ABOUT USER HOME SEARCH CURRENT ARCHIVES

Home > User > Author > Submissions > New Submission

**Step 1. Starting the Submission**

1. START 2. ENTER METADATA 3. UPLOAD SUBMISSION 4. UPLOAD SUPPLEMENTARY FILES 5. CONFIRMATION

Encountering difficulties? Contact Till Dembeck for assistance.

**Journal Section**

Select the appropriate section for this submission (see Sections and Policies in [About the Journal](#)).

Section\*

**Journal's Privacy Statement**

**Comments for the Editor**

Enter text (optional)

**OPEN JOURNAL SYSTEMS**

[Journal Help](#)

**USER**

You are logged in as... **sd**

- [My Journals](#)
- [My Profile](#)
- [Log Out](#)

**JOURNAL CONTENT**

Search

**Browse**

- [By Issue](#)
- [By Author](#)
- [By Title](#)

Abb. 4: Beispiel für die Nutzung eines Peer-Review-Systems



Abb. 5: Virtuelles Blättern in einem digitalisierten Band

## Autorensupport

Der edoc-Server setzt auf Creative Commons Lizenzen [7], um die Nutzungsrechte an digitalen Dokumenten zu verankern. So kann jeder Autor während der Abgabe seiner Dokumente entscheiden, welche Rechte er für das jeweilige Dokument vergibt.

Für Autoren, die sich und ihre Publikation über das Online-Meldesystem der VG Wort angemeldet haben, kann auf dem edoc-Server die entsprechende technische Verankerung für die Zählpixel realisiert werden [8].

Bei Bedarf werden auch komplette Publikationstechnologien, inklusive Dokumentvorlagen, Importschnittstellen aus gängigen Autorensystemen und Exportschnittstellen nach PDF, PDF/A und XML für die unterschiedlichsten Textsatzsysteme, vorrangig Adobe Indesign, Adobe FrameMaker, Corel Draw, aber auch für LaTeX und weitere Systeme erstellt. Autorenvorlagen sind für Microsoft Office, Staroffice und Openoffice sowie LaTeX möglich. Dazu können entsprechende Anwenderdokumentationen oder auch Herausgeber- und Autorenschulungen durchgeführt werden.

## Bibliothekarische Betreuung

Der Betrieb eines edoc-Servers setzt auch die Durchführung entsprechender bibliothekarischer Tätigkeiten voraus. Bei einem edoc-Hosting können die Katalogisierung, die Formalerschließung und die Sacherschließung nach RVK, mit DDC Subject

Headings und die Schlagwortvergabe anhand der Schlagwortnormdatei (SWD) übernommen werden. Auch die Bearbeitung und Vorbereitung der Dokumente sowie die Konvertierung aus Textverarbeitungssystemen in publikationsfähige Formate wie XML oder PDF kann beauftragt werden.

## Datenspeicherung und Datensicherung

Die Ausfallsicherheit des Servers wird durch die Nutzung eines RAID5-Systems und eine SAN (Storage Area Networks) gewährleistet. Es liegen demnach immer mehrere Kopien der Daten vor, sodass bei einem Ausfall einer Festplattenkomponente innerhalb kürzester Zeit (maximal 24 Stunden) die Umschaltung auf eine anderen Plattenspeicherplatz erfolgen kann. Ein Server wird eingebunden in das tägliche Band-Backup über das Tivoli Storage Management System. Die redundante Datenspeicherung erfolgt an den zwei Standorten Berlin-Mitte und Berlin-Adlershof. Es kann auch ein eigenes LOCKSS<sup>1</sup>-Netzwerk bereitgestellt werden.

## Hosting von e-Journals

Auf Basis der edoc-Servertechnologie wird das Hosting von E-Journals angeboten. Die Kunden sind dabei vor allem kleinere bis mittlere wissenschaftliche

<sup>1</sup> LOCKSS: Lots Of Copies Keep Stuff Safe – eine Entwicklung und Initiative zur Bitstromsicherung unter der Federführung der Stanford-Universität – <http://www.lockss.org>, vgl. [9].

Institutionen. Für diese wird die Bereitstellung der technischen Komponente für ein E-Journal inklusive eines kundenspezifischen Metadatenformats und Layouts sowie die Konfiguration eines Begutachtungsworkflows übernommen. Wir bereiten ein medienneutrales Publikationsverfahren mittels XML-Technologie vor. Dazu wird gemeinsam mit den Herausgebern die komplette Publikationstechnologie, angefangen bei der Autoren-Dokumentvorlage, über die Exportschnittstelle nach XML und die XSLT-Umwandlung, bis hin zur Präsentation auf dem edoc-Server angepasst und bereitgestellt.

Bei Bedarf kann auch die bibliothekarische oder die Autoren-Betreuung übernommen werden.

## Literatur

- [1] HUMBOLDT-INNOVATION GMBH: <http://www.humboldt-innovation.de>, (23.03.2009)
- [2] SCHROEDER, K.: Zum Nutzen von „Persistent Identifiers“: von CARMEN-AP4 über EPICUR zum Normalbetrieb. ZfBB Heft 3/4, 2005, S. 181–187
- [3] ARBEITSGRUPPE ELEKTRONISCHES PUBLIZIEREN DINI: DINI-Zertifikat für Dokumenten- und Publikationsservices 2007. April 2007, DINI-Schriften 3, urn:nbn:de:kobv:11-10079197 (18.03.2009)
- [4] <http://pkp.sfu.ca/?q=ojs>, (18.03.2009)
- [5] EPRINTS APPLICATION PROFILE: <http://www.ukoln.ac.uk/repositories/digirep/index/SWAP>, (23.03.2009)
- [6] <http://www.ub.hu-berlin.de/service/eod/eod.html>, (23.03.2009)
- [7] KLATT-KAFEMANN, M.: Creative Commons. CMS-Journal 32, 2009, S. 81–86
- [8] [http://www.vgwort.de/metis\\_news.php](http://www.vgwort.de/metis_news.php), (23.03.2009)
- [9] REICH, V.; ROSENTHAL, D. S. H.; ROBERTSON, T.; LIPKIS, T.; MORABITO, S.: Requirements for Digital Preservation Systems: A Bottom-Up Approach. D-Lib Magazine, November 2005, Volume 11 Number 11, doi:10.1045/november2005-rosenthal, <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>, (20.02.2009)

# EoD bringt alte Bücher nach Hause

Patricia Fechner | patricia.fechner@cms.hu-berlin.de

Wer schon immer einmal die Farbenlehre von Goethe studieren oder an den naturwissenschaftlichen Erkenntnissen der Javareise von Franz Wilhelm Junguhn im Jahr 1845 teilhaben wollte, kann die Bücher nun digital direkt nach Hause bestellen.

Diesen neuen Service macht die Universitätsbibliothek der HU möglich. Er ist das Ergebnis des EU-Projektes „EoD – eBooks on Demand“ mit 13 teilnehmenden Universitäts- und Nationalbibliothe-

vor Ort gestattet wird. Diese Einschränkungen erschweren natürlich vor allem die Arbeit der Wissenschaftler. Mit „EoD – eBooks on Demand“ ändert sich die Situation schlagartig – jetzt hat jeder Zugriff auf diese Wissensschätze, weltweit und zu jeder Zeit.

Der Service ist im Online-Katalog der Universitätsbibliothek eingebettet. Für urheberrechtsfreie Werke, die vor 1909 erschienen sind, wird in der Exemplaransicht der EoD-Button angezeigt.

The screenshot shows the 'HU Berlin Katalog' interface. At the top, there are navigation links: 'Anmelden', 'Neustart', 'Suche', 'Suchgeschichte', 'HU-Kataloge', 'Kontakt', and 'Hilfe/FAQ'. Below this is a search bar and a 'Zurück zur Titelliste' / 'Zurück zur vorigen Seite' link. The main section is titled 'HU Berlin Katalog - Bestand' and displays details for a specific item: 'Zur Farbenlehre / von Goethe. - Wien : Geistinger. Erklärung der zu Goethe's Farbenlehre gehörigen Tafeln, [1812] - 24 S., XVI Bl., 12 S.'. Below the item details, there are filters for 'Wählen Sie ein Jahr', 'Wählen Sie einen Band', and 'Wählen Sie eine Zweigstelle'. A table below shows the 'Ausleihmöglichkeit' (loan possibility) for the item, with columns for 'ausgegeben bis', 'Zweigstelle', 'Standort', 'Signatur 2', 'Signatur', 'Band/Jahr/Heft', and 'Exlibris'. The 'Ausleihmöglichkeit' is 'nur Lesesaal-4W'. The 'Zweigstelle' is 'Zentralbibliothek'. The 'Standort' is 'Geschlossenes Magazin / Rara'. The 'Signatur 2' is '2942:F4'. The 'Exlibris' is 'Kaiser-Wilhe'. There is an 'eod' button in the bottom left corner of the table area.

Abb. 1: Integration des Service „EoD-eBooks on Demand“ im Online – Katalog

“EoD – eBooks on Demand” ist ein neuer Service der Universitätsbibliothek zur Digitalisierung urheberrechtsfreier Werke auf Bestellung und gegen Bezahlung. Ausgangspunkt war ein EU-Projekt mit 13 teilnehmenden Universitäts- und Nationalbibliotheken aus ganz Europa. Seit Beendigung der Projektphase wird dieser Service als Basisdienstleistung angeboten.

EoD; eBooks on Demand; Digitalisierung; urheberrechtsfrei; Service; Universitätsbibliothek

ken aus ganz Europa. Ziel ist die Digitalisierung urheberrechtsfreier Werke auf Bestellung und gegen Bezahlung, um diese Werke für Forschung und Lehre, aber auch für den Privatgebrauch nutzbar zu machen. Diese alten Bücher können, sofern es ihr Zustand erlaubt, nur im Lesesaal gelesen werden. Eine Ausleihe oder das Kopieren einzelner Seiten ist aus Bestandsschutzgründen nicht gestattet. Oft sind weite Reisewege in Kauf zu nehmen, um die Werke anzusehen. Einige Bücher oder Manuskripte sind so alt, dass noch nicht einmal das Lesen

Mit diesem Button gelangt der interessierte Leser direkt auf die Online-Bestellseiten und kann die Digitalisierung des Buches in Auftrag geben. Die Bestellung ist sehr einfach und vergleichbar mit anderen Webportalen. Im ersten Schritt wird eine Übersicht mit den Buchdaten und dem Preis für die Digitalisierung angezeigt. Über die Links im Servicebereich der Bestellseiten können sich die Kunden jederzeit über unseren Service, die Allgemeinen Geschäftsbedingungen, die Möglichkeiten der Bezahlung und die Lieferfristen informieren.



BESTELLUNG ÜBER EOD

Ihr gewünschter Buchtitel

Der gewünschte Titel ist **noch nicht digital verfügbar**. Geben Sie als erste die Digitalisierung dieses Buches in Auftrag!

Autorinnen

Reihentitel

Titel

Band

Jahr

Anzahl Seiten

Verlag

Signatur

Preis (Standard eBook) (€)

Erklärung der zu Goethe's Farbenlehre gehörigen Tafeln

1812

36

12.5

1 eBook bestellen

Bestellen Sie jetzt dieses Buch als eBook!


Sie können bei der Bestellung zusätzliche Optionen (z.B. Lieferzeit) auswählen.

eBook jetzt bestellen >

Kundenservice

☒ EOD: Das Angebot im Überblick
☒ Bezahlung & Lieferung
☒ AGB
☒ eBooks on Demand
☒ FAQ
☒ Kontakt

Deutsch



Universitätsbibliothek der Humboldt-Universität zu Berlin

Abb. 2: Anzeige der Buchdaten im ersten Schritt

BESTELLUNG ÜBER EOD

Gewünschten Buchtitel als eBook bestellen:

Erklärung der zu Goethe's Farbenlehre gehörigen Tafeln, 1812

Preis (€): 12.5

Buch direkt bestellen. Hinweis: Diese Bestellung verpflichtet zum Kauf.

Geben Sie bitte Ihre persönlichen Daten ein:

Mit \*) gekennzeichnete Felder sind Pflichtfelder

Als Kunde registrieren

Warum registrieren?

Vorname

Nachname

Anrede

Straße/Hausnr.

PLZ/Stadt

Land

US-Bundesstaat

E-Mail

Telefon

Institution/Abteilung

Bibl. Ausw. ltr.

Firmenrechnung

Lieferung

Bemerkung

Klaus

Mustermann

Frau

Musterstrasse 12

11111

Musterstadt

Deutschland

Außerhalb USA

klaus.mustermann@gmx.net

Download

☒ Ich habe die AGB gelesen und erkläre mich damit einverstanden. \*)


Preis neu berechnen >

Auftrag fertig stellen >

Kundenservice

☒ EOD: Das Angebot im Überblick
☒ Bezahlung & Lieferung
☒ AGB
☒ eBooks on Demand
☒ FAQ
☒ Kontakt

Deutsch



Universitätsbibliothek der Humboldt-Universität zu Berlin

Abb. 3: Eingabe der Kundendaten

Ist der Leser mit unseren Konditionen einverstanden und möchte das Buch zu diesem Preis als eBook bestellen, müssen nun die Kundendaten eingegeben werden. Der Kunde kann sich am Bestellsystem registrieren. In diesem Fall erhält der Nutzer seine Zugangsdaten per Email und die Eingabe der Kundendaten entfällt bei allen weiteren Bestellungen.

Sobald die Bestellung im System gespeichert wurde, erhält der Kunde eine E-Mail mit allen relevanten Informationen, wie z. B. die Buchdaten, die Lieferadresse, den Preis und die Lieferzeit. Mit dieser Email wird auch der Link zur persönlichen Kundenseite verschickt. Über diese Seite kann man sich jederzeit den Status der Auftragsbearbeitung ansehen, online bezahlen sowie das fertige eBook herunterladen. Die vereinbarte Lieferfrist beträgt 15 Werktage. Zurzeit werden die Aufträge in der Regel innerhalb von 7 Werktagen abgearbeitet und die eBooks können nach erfolgter Bezahlung an den Leser ausgeliefert werden.

Reine Textseiten werden in Schwarz-Weiß mit hoher Qualität, standardmäßig 400 dpi, gescannt; Bilder, Illustrationen sowie Fotos je nach Vorlage in Graustufen oder farbig. Mit der Software zur automatischen Texterkennung (OCR) wird das eBook anschließend für die Volltextsuche aufbereitet. So kann gezielt nach Schlagwörtern gesucht werden, sogar bei Werken in sehr alter Frakturschrift.

Der Nutzer kann sein eBook am Bildschirm lesen, blättern und zoomen sowie das gesamte Buch oder ausgewählte Seiten ausdrucken. Texte und Bilder können ausgeschnitten und in eigenen Arbeiten als Zitate eingefügt werden.

Ausgeliefert wird das eBook über einen Downloadlink, auf Wunsch aber auch auf CD. Der Preis für die Digitalisierung beträgt 0,20 € / Seite. Ein 100-seitiges Werk kostet 20,00 €; für 200 Seiten bezahlen interessierte Leser 40,00 €. Der Mindestpreis pro Werk beträgt 12,50 €. Für die Lieferung auf CD fallen dann noch Kosten laut Gebührenordnung der UB von 8,50 € für Europa und 10,50 € für den weltweiten Versand an. Beahlt wird das eBook online mit Kreditkarte oder per Click&Buy; für ausgewählte Bestände bietet die Zentralbibliothek die Möglichkeit der Vor-Ort-Bezahlung.



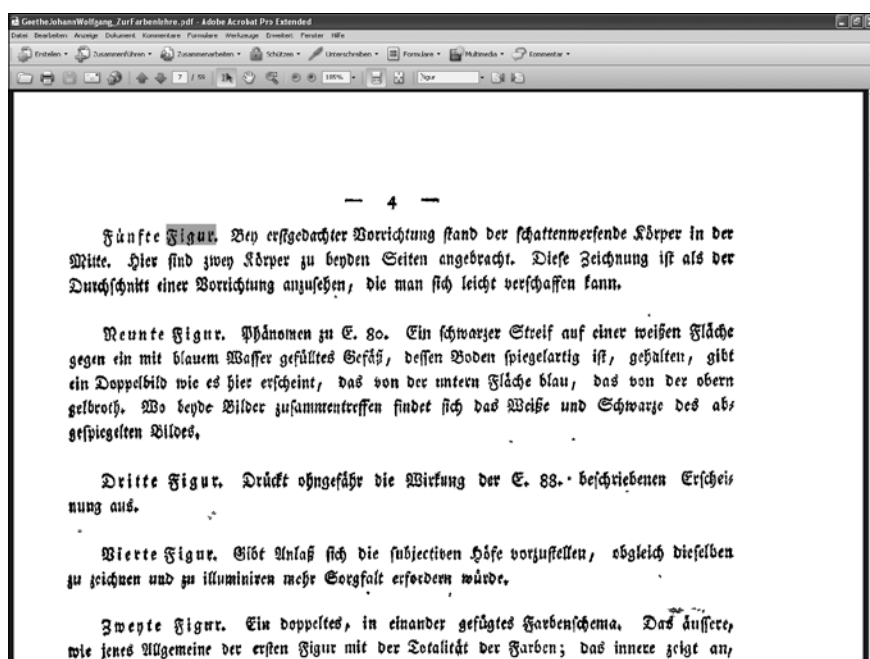


Abb. 4: Volltextsuche im eBook

## Fazit

Nun, ein Jahr nach Start des neuen Services, können wir eine durchaus positive Bilanz ziehen. Nach anfänglich sehr verhaltener Auftragslage stiegen die Aufträge ab Januar 2008 sprunghaft an. Wir haben dabei nicht nur einzelne Bücher für den privaten Gebrauch digitalisiert. Erfreulich ist auch die Tatsache, dass wir Institutionen, wie die Berlin-Brandenburgische Akademie der Wissenschaften oder das Museum für Naturkunde, als Kunden gewinnen und so die Forschungsarbeit dieser Einrichtungen maßgeblich unterstützen konnten.

Nach Beendigung der Förderung durch die Europäische Union im Juni 2008 nahm die Frage nach der Weiterführung des Services bei allen Projektpartnern natürlich eine vordergründige Rolle ein. Unter Federführung von Innsbruck wurde ein Konsortium mit den am Projekt beteiligten Bibliotheken gegründet. Zudem ist es gelungen, weitere Einrichtungen, wie die Sächsische Landes- und Universitätsbibliothek Dresden, die Moravian Library in Brno und die Medical and Dental Academic Library in Paris, für eine Mitarbeit im Konsortium zu gewinnen.

Alle Mitglieder leisten einen finanziellen Beitrag zur Nutzung der technischen Infrastruktur in Innsbruck, ins-

besondere der OCR-Software und gewährleisten so die Nachhaltigkeit des Services.

Die EoD-Administration innerhalb der Humboldt-Universität wurde mit Beendigung der Projektphase vom Computer- und Medienservice an die Universitätsbibliothek übergeben. Diese Aufgabe wird, erst einmal befristet für 1 Jahr, von Frau Hühn aus der Zweigbibliothek Naturwissenschaften übernommen. Um den Service auch in Zukunft in der gewohnt hohen Qualität anbieten zu können, muss hier über eine dauerhafte Lösung von Seiten der Universitätsbibliothek nachgedacht werden.

## Weiterführende Informationen unter:

EoD-Administrator  
Universitätsbibliothek der HU  
Zentralbibliothek  
Referat Historische Buchbestände

Hessische Straße 1-2  
10115 Berlin

Tel: 2093-6920  
Email: [eod.admin@cms.hu-berlin.de](mailto:eod.admin@cms.hu-berlin.de)  
<http://books2ebooks.eu/>

# Grundfragen der digitalen Langzeitarchivierung für den edoc-Server

Susanne Dobratz | [dobratz@cms.hu-berlin.de](mailto:dobratz@cms.hu-berlin.de)

Digital Information lasts forever or five years – whichever comes first.  
Jeff Rothenberg [1]

## Der edoc-Server der Humboldt-Universität zu Berlin

Seit 1998 ist der edoc-Server an der Humboldt-Universität in Betrieb. In den vergangenen 11 Jahren haben wir als Betreiber des Servers so einige Technologiewechsel miterlebt. Wir haben die Hardware des edoc-Servers mehrfach erneuert und entsprechend dazu neue Betriebssystemversionen des Sun Solaris-Betriebssystems installiert. Wir haben den Kern unseres edoc-Servers, die Web-Anwendungen, ausgebaut. Vor allem aber gab es seitdem einige Versionswechsel des Textverarbeitungssystems Microsoft Word, mit dem der Großteil der Autoren seine Publikationen erstellt. Diese haben dazu geführt, dass die Dokumentvorlagen, besonders die für die Dissertationen, mehrfach angepasst werden mussten. Auch das Publikationsformat PDF hat seit 1998 mehrere Versionswechsel erlebt.

Wagt man unter diesen Bedingungen einen Blick in die Zukunft, so ist eines sicher: Technologiewechsel finden im IT-Bereich sowohl bei der Hardware als auch bei der Software relativ schnell statt. Man geht davon aus, dass alle 2–3 Jahre eine Umstellung einzelner Komponenten erfolgen muss. Diesen stetigen Wandel an Technologie zu bewältigen und den edoc-Server als stabile Dienstleistung bereitzustellen stellt eine große Herausforderung dar. Denn das Ziel ist es, den edoc-Server als zentrale Publikationsplattform für die Universität auch in den nächsten Jahrzehnten zu betreiben.

So heißt es in den Leitlinien des edoc-Servers: „Der Dokumenten- und Publikationsserver bietet allen Angehörigen der Humboldt-Universität die organisatorischen und technischen Rahmenbedingungen zur elektronischen Publikation wissenschaftlicher Dokumente. Im Rahmen dieses Gemeinschaftsangebotes von Computer- und Medienservice und Universitätsbibliothek der Humboldt-Universität werden wissenschaftliche Dokumente von hoher Relevanz unter Einhaltung von Qualitätsstandards im Internet für Forschung und Lehre bereitgestellt. Die elektronischen Dokumente erhalten dauerhafte Adressen und werden über nationale und internationale Bibliothekskataloge, Suchmaschinen sowie andere Nachweisinstrumente erschlossen. Der Dokumenten- und Publikationsserver bietet durch besondere Maßnahmen wie digitale Signaturen und Zeitstempel einen Schutz gegen Verfälschungen. Darüber hinaus wird eine Langzeitarchivierung der elektronischen Dokumente gewährleistet.“ [2]

Die Problematik der Langzeitarchivierung wurde beim Aufbau des Dokumentenservers der HU von vornherein als ein wichtiges Thema angesehen, und die dazu erforderlichen Maßnahmen wurden und werden schrittweise umgesetzt.

*In diesem Artikel wird das LZA-Konzept „edoc-Server“ vorgestellt. Dabei werden verschiedene Methoden beleuchtet, angefangen bei dem Einsatz archivierungsfähiger Dateiformate, bis hin zur Herstellung technischer Redundanz über RAID-Verfahren.*

**Keywords**

## Was versteht man unter digitaler Langzeitarchivierung?

Das Ziel der Archivierung digitaler Dokumente ist es, diese für eine zukünftige Nutzung so aufzubewahren, dass kommende Generationen darauf zugreifen, sie benutzen und Zugang zu dem darin kodierten intellektuellen und kulturellen Inhalt finden können. Idealerweise würde man davon ausgehen, dass jedes digitale Dokument in seiner heutigen Form und Funktionsweise verfügbar und benutzbar gehalten wird. Allerdings ist dieses Ideal kaum zu realisieren, da die Technologien sich sehr schnell und mitunter auch sehr stark verändern. Deshalb gehen heutige Strategien zur digitalen Archivierung davon aus, dass es unmöglich ist, zukünftige Technologien und Anwendungsszenarien soweit im Voraus zu bestimmen. Daher ist es im Allgemeinen das vorrangige Ziel, zum einen den Zugriff auf die digitalen Dokumente sicher zu stellen und zum anderen deren Lesbarkeit.

Wenn man z. B. ein eigenes Dokument heute in digitaler Form auf einem universitären Dokumentenserver bereitstellt, möchte man natürlich auch, dass die Enkel und Urenkel dieses Werk noch lesen und benutzen können. Dies gelingt allerdings nur, wenn die Universität dafür sorgt, dass entsprechende Maßnahmen für die Langzeitarchivierung der digitalen Publikation getroffen werden. Digitale Langzeitarchivierung beginnt beim Produktionsprozess. Als Autor schreibt man meist mit dem am besten verfügbaren Textverarbeitungsprogramm, etwa Microsoft Office oder auch OpenOffice. Dabei stellt sich die Frage, ob diese Dateien in drei oder fünf Jahren noch in dem Nachfolgeprogramm geöffnet werden können. Viele Computernutzer haben die Programme, mit denen die Dokumente vor einigen Jahren geschrieben worden sind, bereits heute nicht mehr auf ihrem Arbeitsplatzrechner verfügbar. Dokumente müssen deshalb in einem standardisierten Dateiformat publiziert werden, etwa in PDF oder XML. Aber auch diese Formate werden möglicherweise im Laufe der Zeit unbenutzbar, sodass die Anwendung einer Archivierungsstrategie ratsam ist, die eine regelmäßige Migration der Dokumente in

aktuelle Dokumentformate vorsieht. An dieser Stelle fällt der Blick auf das Speichermedium, das benutzt wurde. Auch die DVD oder magnetische Medien wie Festplatten altern physikalisch. Demnach muss die Speicherhardware auch regelmäßig überprüft und erneuert werden. Die Identifizierung des zum betrachteten Dokument gehörenden Bitstroms und die Identifikation und Interpretation des Dateiformates spielen eine wichtige Rolle für dessen spätere Benutzbarkeit. Dazu kommt die Bereitstellung des Daten- oder Zusatzmaterials, der Statistiken oder der Videos, die zum Inhalt heutiger wissenschaftlicher Arbeiten gehören. Welches aber sind die Standards, die wir heute schon einsetzen können, um zukünftigen Generationen den Zugang und die Benutzbarkeit zu unserem digitalen Kulturerbe zu garantieren? Der schnelle Technologiewechsel unserer Zeit macht es uns unmöglich, diese Art von Garantie zu geben. Im Gegenteil, wir werden permanent damit beschäftigt sein, diesen Technologiewechsel zu beobachten und angemessen darauf zu reagieren.

„Langzeitarchivierung ist nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.“ [3]

Dabei spielen sowohl organisatorische als auch technische Maßnahmen eine große Rolle, wie sie u. a. im „Kriterienkatalog vertrauenswürdige digitale Langzeitarchive“ [4] formuliert werden. Sie schließen sowohl die Migration als auch die Emulation als Erhaltungsmethode ein.

Unter Migration wird eine Menge organisierter Abläufe verstanden, die den regelmäßigen Transfer digitalen Materials von einer Hardware/Software-Konfiguration in eine andere realisiert oder von einer Computer-Technologie-Generation in eine nachfolgende Generation. Dagegen versteht man unter Emulation die Methode, ein digitales Dokument (mit seiner originalen Kodierung) unter Nutzung einer speziellen Software (eines Emulators) in seiner ursprünglichen Softwareumgebung nutzbar zu machen.

## Organisatorische Rahmenbedingungen

Langfristig soll der edoc-Server zu einem vertrauenswürdigen Langzeitarchiv ausgebaut werden.

Digitale Langzeitarchive haben den Erhalt der Informationen über lange Zeiträume hinweg zum Ziel. Deshalb ergreifen sie sowohl organisatorische als auch technische Maßnahmen, um den Bedrohungen, denen die digitalen Objekte ausgesetzt sind, entgegenzuwirken. Dazu gehört die Nutzung von Standards in Bezug auf die Hardware, die Speicherformate der digitalen Objekte, die Beschreibung mit angemessenen Metadaten sowie der Einsatz standardisierter Verfahrensweisen bei der Migration oder der Emulation digitaler Objekte. Vertrauenswürdige digitale Langzeitarchive operieren nach ihren Zielen und Spezifikationen, vgl. [4].

Bis dies realisiert ist, ist die Humboldt-Universität zu Berlin eine Kooperation mit der Deutschen Nationalbibliothek (DNB) eingegangen. Die über den edoc-Server publizierten digitalen Dissertationen und Habilitationsschriften werden in das DNB-eigene digitale Langzeitarchiv übernommen.

Einige Eigenschaften eines vertrauenswürdigen digitalen Langzeitarchivs besitzt der edoc-Server aber bereits schon heute? Dazu gehört an erster Stelle die Garantie eines zuverlässigen und organisatorisch gesicherten Betriebes. Diese Garantie hat die Humboldt-Universität zu Berlin dadurch gegeben, dass sie den edoc-Server als festen Bestandteil in das Dienstleistungsangebot von Computer- und Medienservice und Universitätsbibliothek eingegliedert und den Betrieb durch zugehöriges Personal abgesichert hat.

An zweiter Stelle wurden für den edoc-Server Leitlinien definiert, die beschreiben, welche Publikationen unter welchen Bedingungen über den edoc-Server veröffentlicht werden und unter welchen Bedingungen Garantien für eine zukünftige Lesbarkeit und Benutzbarkeit der Dokumente gegeben werden.

Drittens werden der Zugang zum und der Zugriff auf den edoc-Server über diverse organisatorische und technische

Maßnahmen kontrolliert und gesichert. So steht der Server selbst im Rechneraum des Computer- und Medienservice und ist damit in die Sicherheits-Infrastruktur des Computer- und Medienservice eingebunden. Durch den Einsatz von digitalen Signaturen wird sichergestellt, dass die Authentizität und Integrität der Dokumente gewahrt wird und jede nachträgliche Änderung der Autorenschaft, des Veröffentlichungsdatums oder gar des Inhaltes bemerkt wird. Des Weiteren wird durch die edoc-spezifische Backupstrategie dem Verlust von Dokumenten vorgebeugt.

Ein vertrauenswürdigen digitales Langzeitarchiv bedient sich transparenter Technologien und Methoden zur Speicherung, Migration und Bereitstellung von Dokumenten und Metadaten. Daher arbeitet die AG Elektronisches Publizieren stetig daran, eine umfassende Dokumentation des edoc-Servers mit sämtlichen angewendeten Verfahren vorzuhalten.

## Konkrete Ansätze

Für den edoc-Server der Humboldt-Universität werden mehrere Strategien verfolgt. Den Ausgangspunkt dieser Überlegungen bildet die Tatsache, dass digitalen Publikationen von unterschiedlichen Seiten betrachtet werden können:

1. Zum einen sind digitale Dokumente nichts weiter als Bitströme, die es zu erhalten und zu sichern gilt.
2. Zum anderen sind digitale Dokumente auch kodierte Informationseinheiten, die in unterschiedlichen Dateiformaten vorliegen können.
3. Zum dritten sind digitale Dokumente auch immer Informationsobjekte, die sehr komplex sein können, z. B. dadurch, dass sie aus mehreren Teilobjekten und Medienelementen bestehen können.

### Backup/ Replikation und Bitstream-Preservation

Die Grundlage für alle weiteren Maßnahmen bildet daher das Konzept zum Backup und zur Replikation der Bitströme. Mit Hilfe eines SAN (Storage Area Networks) zur redundanten Spei-

cherung mehrerer Kopien der Daten an unterschiedlichen Standorten sowie dem täglichen Backup und der täglichen Kopie der Daten wird die Existenz des Bitstroms gesichert, siehe dazu die ausführlichere Beschreibung in [5].

Zusätzlich dazu betreibt die Arbeitsgruppe Elektronisches Publizieren auch einen LOCKSS-Server und beabsichtigt, einen ausgewählten Teil des Inhaltes des edoc-Servers dadurch in weltweite Netze zu replizieren. Der Einsatz dieser Technologie ist vor allem für die über den edoc-Server veröffentlichten e-Journals geplant.

LOCKSS, *Lots Of Copies Keep Stuff Safe*, wurde ursprünglich in einem Projekt der Universität Stanford entwickelt, vgl. [6]. Es ist ein so genanntes Peer-to-Peer-Netzwerk, mit dessen Hilfe Daten an mehreren Standorten verteilt und beobachtet werden. Bei Bedarf werden kaputte Bitströme durch einen Replikationsalgorithmus repariert. Ziel ist es, die Unversehrtheit von durch Bibliotheken abonnierten E-Journals auf der Bitebene zu garantieren. Dabei stellt das eigentliche LOCKSS-System die technische Infrastruktur zur Verfügung, um einen Datenverlust auf physikalischer Ebene zu verhindern. Ein Datenverlust tritt zum Beispiel ein, wenn E-Journals gestrichen werden, wenn Bibliotheken ihre Subskriptionen beenden müssen, sobald ein E-Journal oder ein Verlag an einen neuen Rechteinhaber verkauft wird, Webseiten von Verlagen nicht mehr erreichbar sind oder auch wenn Open-Access-Literatur aus dem WWW verschwindet. Die LOCKSS-Technologie setzt auf „Low-Tech“, d. h. es wird eine Open-Source-Software (besonderes OpenBSD, seit 1998), deren komplette Dokumentation frei zugänglich ist, genutzt. Die LOCKSS-Teilnehmer müssen eine LOCKSS-Box aufsetzen. Über das Peer-to-Peer-Netzwerk werden im Sechswochen-Takt neue Versionen der LOCKSS-Software automatisch eingespielt. Des Weiteren werden durch die Stanford-Universität automatische Tools zum Nutzen der Software (z. B. zur Erstellung von Plug-Ins) bereitgestellt. Die LOCKSS-Box entspricht einem Computer, auf dem die LOCKSS-Software läuft. Sie ist über ein Web-User-Interface konfigurier- und administrierbar und sammelt

die Inhalte von Verlagsseiten, via crawl über http. Dabei können alle Formate: HTML, PDF, JPEG, TIF, Audio, Video, Excel, Java genutzt werden. Gesammelt wird die Präsentationsform der Inhalte. Eine LOCKSS-Box arbeitet als Web-Proxy, um den eigenen Inhalt anzubieten und gleicht sich über einen Peer-to-Peer-Prozess mit den anderen LOCKSS-Boxen ab. Dabei werden notfalls beschädigte Daten durch den speziellen LOCKSS-Algorithmus repariert.

### Erhaltung der Benutzbarkeit und Lesbarkeit von Dokumenten – Erhaltung auf der logischen Ebene – der Ebene der Dateiformate

Dateiformate sind wohl der Faktor bei der digitalen Langzeitarchivierung, der am prägnantesten dem technischen Wandel und der Obsoleszenz unterworfen sind. Die Anzahl der heute existierenden Dateiformate ist kaum abzuschätzen, da für jede erdenkliche Anwendung bzw. jedes mögliche Anwendungsprogramm eigene Formate entwickelt werden können. Für die Langzeitarchivierung wird als archivierbare Einheit das digitale Dokument angesehen. An dieser Stelle wird in Anlehnung an [7] als Dokument eine geordnete und abgeschlossene Einheit von Informationseinheiten verstanden, die in einer festen Struktur organisiert und mit Formatierungsregeln assoziiert ist, die es sowohl den Produzenten als auch Rezipienten gestattet, es zu „lesen“.

Dieses kann aus einer oder mehreren Dateien bestehen. Dabei versteht man unter einer Datei eine geordnete und abgeschlossene Sammlung von Informationseinheiten, die bei der Speicherung als Einheit betrachtet wird. Auf heutigen Systemen ist dies üblicherweise eine beliebige Anzahl eindimensional adressierter Bytes. Das Dateiformat bezeichnet alle Konventionen der inneren Struktur und Anordnung der Bytes, die die Form der Abspeicherung von Computerdateien bestimmen.

Für den edoc-Server existiert die folgende Strategie zum Umgang mit Dateiformaten: Angestrebt wird bei den meisten Publikationsvorhaben die Abspeicherung der Dokumente in einem

archivierungsfähigen Format. Das bedeutet vor allem, dass es bestimmte Eigenschaften besitzt: so muss das Format z. B. dokumentiert und offen gelegt sein, es sollte durch ein internationales, herstellerunabhängiges Gremium normiert bzw. standardisiert worden und auf verschiedenen Betriebssystemplattformen nutzbar sein. Zudem sollte es eine gute Verbreitung haben, damit die Existenz vielfältiger Softwarewerkzeuge zu dessen Verarbeitung gesichert ist. Es sollte eine Dokumentation der Revisionen bzw. eine Rückwärtskompatibilität gegeben sein. Einen wichtigen Aspekt bildet dabei auch die Tatsache, dass das Format in der Lage sein muss, die Struktur und Abfolge der Informationen so zu erfassen, dass diese extrahierbar sind (weitere Aspekte siehe [8]).

Da Markup-Sprachen aus heutiger Sicht diesen Anforderungen am ehesten gerecht werden, wurde die Nutzung XML als Strategie für die Langzeitarchivierung des edoc-Servers gewählt. Für die Publikationskategorien Dissertationen, Habilitationsschriften sowie Diplom- und Masterarbeiten kann diese Strategie zu ca. 85 Prozent umgesetzt werden. Diese Dokumente werden in der Regel von der AG Elektronisches Publizieren (EPUB-Team) nach XML konvertiert und zusätz-

lich zum digitalen Druckexemplar im PDF-Format gespeichert. Sie werden über ein XSLT-Stylesheet formatiert und so auf dem edoc-Server visualisiert.

Der Sinn der Nutzung von XML als Archivformat besteht darin, spätere Migrationen in neue Anzeigeformate einfacher, automatisiert und somit kostengünstiger vornehmen zu können. Die Nutzung von XML zur langfristigen Speicherung von Dokumenten hat sich in den letzten Jahren als geeignete Strategie für digitale Langzeitarchive herausgebildet und wird auch in größeren, renommierteren Projekten, u. a. in PORTICO [9], eingesetzt.

Nicht alle Dokumente des edoc-Servers können nach XML konvertiert werden. Dies hat verschiedene Ursachen:

1. Die Anlieferung der Dokumente durch Autoren erfolgt bereits in einem Format, welches sich nicht ohne Weiteres nach XML konvertieren lässt. Dies ist zum Beispiel bei den Open-Access-Publikationen im edoc-Bereich *Pre- und Postprints* der Fall.
2. Das EPUB-Team verfügt bisher über keine ausgereifte und effiziente Technologie der Umwandlung. So können z. B. LaTeX-Dokumente nur teilweise übertragen werden. Ein Problem hierbei stellt die strukturierte und korrekte

Überführung mathematischer Formeln nach XML dar.

3. Es handelt sich um einen multimedialen Bestandteil eines Dokumentes, wie z. B. eine Anlage in Form kleiner Video- oder Audiosequenzen oder gar um einen kompletten Film.

Bei Multimediainhalten wird die Strategie verfolgt, daß zum aktuellen Zeitpunkt möglichst Formate eingesetzt werden, die einen Standard im Sinne einer Norm darstellen und vor allem keine Komprimierungsalgorithmen beinhalten. Sollte zu einem zukünftigen Zeitpunkt das Format nicht mehr unmittelbar durch einen Nutzer lesbar sein, hofft man, dass es einen geeigneten Emulator für dieses Format geben wird.

So besteht die Hauptstrategie des edoc-Servers bezüglich der Langzeitarchivierung darin, zum einen die Existenz der Bitströme zu sichern und zum anderen die Vielfalt der Formate auf dem edoc-Server auf diejenigen einzuschränken, die am besten für eine Langzeitarchivierung geeignet sind. Für diese können dann auch geeignete Garantien gegeben werden. So heißt es in den Leitlinien des edoc-Servers: „Bei Verwendung des Formates SGML/XML wird eine Archivierungsgarantie von 50 Jahren gegeben. Die Archivierungsdauer anderer Formate hängt von der Verfügbarkeit des Formates, der Betrachtungssoftware sowie den Konvertierungsmöglichkeiten ab.“

Als zweites Dateiformat wird auf die Nutzung von PDF/A orientiert. PDF/A ist ein ISO-Standard (ISO 19005) für langzeitarchivierbares PDF. Er schreibt detailliert vor, welche Kodierungen in PDF-Dokumenten erlaubt sind und welche nicht. In der Norm sind zwei Konformitätsebenen spezifiziert: PDF/A-1a - Level A conformance: Das Dokument muss eindeutig visuell reproduzierbar sein, Text muss nach Unicode abbildbar sein und es muss eine inhaltliche Strukturierung des Dokuments geben. Bei der PDF/A-1b - Level B conformance genügt die eindeutige visuelle Reproduzierbarkeit des Dokumentes. Verboten sind in PDF/A Referenzierungen auf externe Objekte und Ressourcen, da diese zukünftig nicht mehr zugänglich sein könnten. Um Letzteres zu verhindern, müssen in

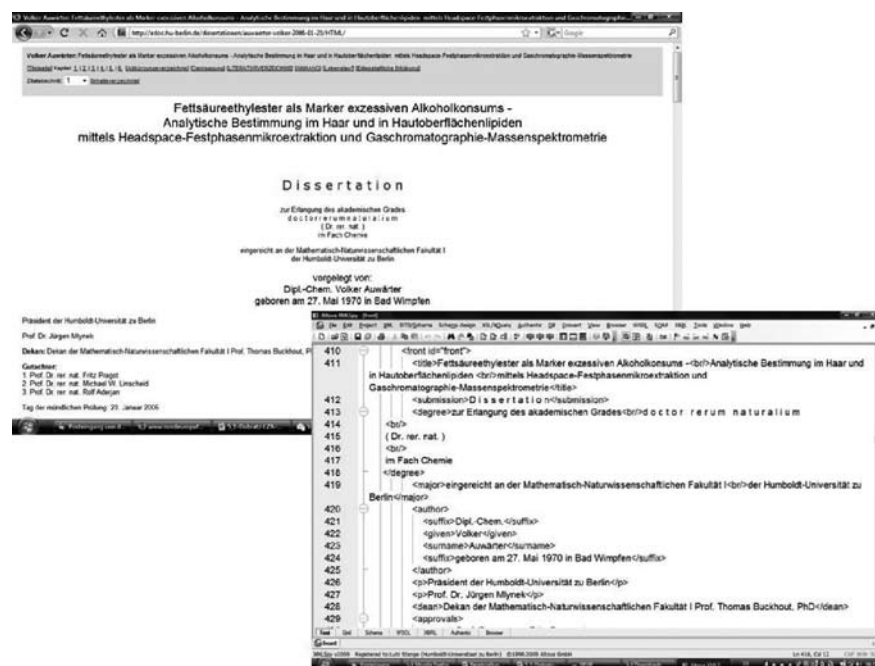


Abb. 1: Ausschnitt aus einem XML-codierten Dokument des edoc-Servers und dessen formatierte und mittels XSLT erzeugte Ansicht



einer PDF/A-1-Datei alle benutzten Schriftarten, alle Bilder, die Kennzeichnung als PDF/A-1 durch Metadaten im Extensible Metadata Platform-Format (XMP) vorhanden sein. Farben müssen, ähnlich wie in PDF/X, ausreichend definiert sein, um eine eindeutige Farbdarstellung zu garantieren. Das entsprechende Quellprofil oder ein „Output Intent“ muss eingebettet werden. Es ist u. a. untersagt, alternative Bilder (z. B. niedrigaufgelöste Varianten für die Bildschirmausgabe und hochaufgelöste Varianten für den Druck) zu integrieren. PDF/A gestattet keine Verschlüsselungen oder das Sperren von Funktionen der Datei wie z. B. das Drucken und das Kopieren von Daten aus der PDF-Datei heraus. Weiterhin ist die Einbettung von Programmiercode, z. B. JavaScript, verboten, da dessen Ausführung den Inhalt oder die Darstellung des Dokumentes verändern kann. Audio- oder Videodaten dürfen nicht in PDF/A-Dokumente integriert sein. Allerdings unterstützt PDF/A die Einbettung von digitalen Signaturen, vgl. [10]. Durch diese und andere Vorschriften soll eine langfristige Benutzbarkeit der Dokumente garantiert werden. PDF/A ist unabhängig von jedweder Anwendungssoftware oder Betriebssystemen.

### Authentizität und Integrität digitaler Dokumente

Im Prinzip ist es für alle edoc-Publikationen wichtig, dauerhaft und rechtlich relevant nachweisen zu können, dass die digitalen Dokumente vollständig und unverändert auf dem edoc-Server verfügbar sind. Dies nennt man Sicherung der Integrität der Dokumente. Zum anderen muss für einen Teil der Dokumente nachweisbar sein, dass der Ersteller und der Veröffentlichungszeitpunkt nachweisbar sind, die auf dem edoc-Server angegeben werden. Dies nennt man Sicherung der Authentizität.

Mit der Erstellung entsprechender Hashwerte sowie dem Anbringen digitaler Signaturen und Zeitstempel an diese können für die edoc-Dokumente, besonders für die Qualifikationsarbeiten, rechtssicher deren Integrität und Authentizität gesichert werden. Die Details zu dem Verfahren sind im Artikel [10] nachzulesen.

Die Nutzung digitaler Hashwerte und Signaturen allein verhindert natürlich nicht, dass Manipulationen an den Dokumenten vorgenommen werden. Hierzu sind weitere organisatorische Sicherheitsmaßnahmen wie Zutritts- und Zugangskontrolle zu den Serverräumen sowie die Zugriffskontrolle durch Vergabe von Benutzerrechten auf das System des edoc-Servers und die darin befindlichen Dokumente erforderlich.

Hashwerte und digitale Signaturen können allerdings dazu dienen, unautorisierte Änderungen an den Dokumenten festzustellen.

### Metadaten und dauerhafte Adressierbarkeit

Für ein langfristiges Management der im digitalen Archiv befindlichen Dokumente benötigt man geeignete Metadaten. Bisher werden für den edoc-Server vorrangig bibliographische und organisatorische Metadaten erfasst. Technische Metadaten werden derzeit nur ausschnittsweise gespeichert. Hier besteht die Notwendigkeit, die Menge der erfassten technischen Metadaten so zu erweitern, dass spezifischere Informationen zu den Dokumentformaten und der Erstellungssoftware in Anlehnung an das METS-Metadatenmodell, kombiniert mit dem Metadatenmodell der PREMIS-Arbeitsgruppe oder alternativ dem LMER-Metadatenmodell, mit aufgenommen werden. Ein Konzept zur Speicherung der Informationspakete

mit den zu jeder Version bzw. jedem Format gehörenden Metadaten wurde bereits erarbeitet und befindet sich derzeit in der Umsetzungsphase, siehe dazu [11]. Dabei wird beim edoc-Server der Humboldt-Universität zu Berlin ein besonderer Wert auf die Einbindung digitaler Signaturen in ein Metadatenkonzept gelegt.

Ein für die Langzeitarchivierung entscheidender Parameter ist die Vergabe konsistenter und dauerhafter Adressen, um die Wiederauffindbarkeit der Dokumente zu gewährleisten. Für alle Dokumente auf dem edoc-Server der Humboldt-Universität werden persistente Identifikatoren in Form der URN (Uniform Resource Name) vergeben. Dabei wird auf das Konzept der National Bibliographic Number (nbn) der Deutschen Nationalbibliothek zurückgegriffen, vgl. [12].

So wird zum Beispiel für das Dokument mit der URL <http://edoc.hu-berlin.de/series/nestor-materialien/2008-8/PDF/8.pdf>, folgende URN vergeben: `urn:nbn:de:0008-2008021802`. Sowohl der edoc-Server als auch der Resolverdienst bei der Deutschen Nationalbibliothek führen eine Datenbank, in der beide Zeichenketten einander zugeordnet werden. Ändert sich zu einem späteren Zeitpunkt der physische Speicherort des Dokumentes z. B. auf [future1.cms.hu-berlin.de](http://future1.cms.hu-berlin.de) wird in der Resolverdatenbank die der URN zugeordnete URL aktualisiert und der Nutzer kann trotzdem weiter mit der alten URN auf das nun an anderer Stelle gespeicherte Dokument referenzieren.

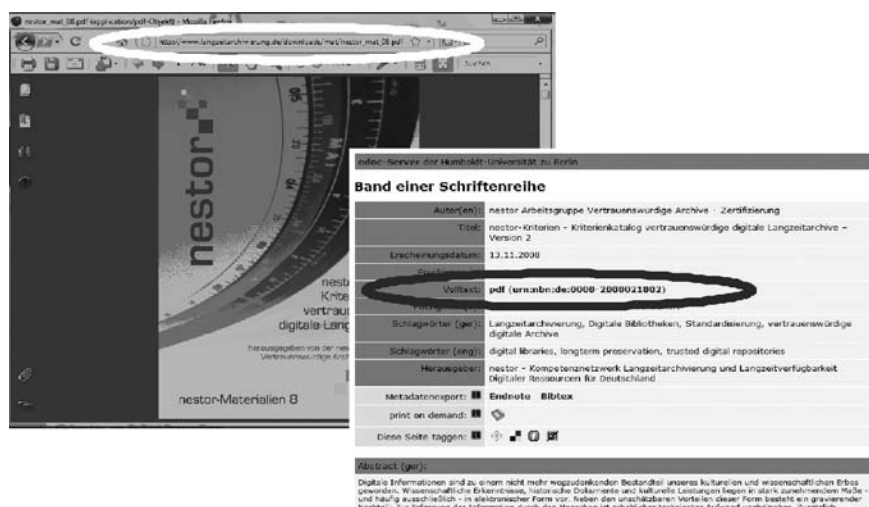


Abb. 2: Beispiel für eine URN in Form der nbn auf dem edoc-Server (der Aufruf des Dokumentes erfolgt über: <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>)

## Pflichtabgabegesetz

Langzeitarchivierung ist eine nationale und Institutionen übergreifende Aufgabe. Dies ist ein Grund für das seit Juni 2006 existierende „Gesetz über die Deutsche Nationalbibliothek“. In diesem ist auch die Pflichtabgabe elektronischer Dokumente, wie sie z. B. hier über den edoc-Server publiziert werden, geregelt. Eine automatische Bereitstellung der Informationen zu den neuen, auf dem edoc-Server veröffentlichten Dokumenten erfolgt über die Bereitstellung der URN-Informationen über die OAI-Schnittstelle des edoc-Servers. Diese kann von der Deutschen Nationalbibliothek abgefragt und die Dokumente und deren Metadaten können so direkt abgerufen werden. Zurzeit ist ein vollautomatisiertes Melde- und Abgabeverfahren für Dokumente an die DNB zwar noch nicht im Produktivbetrieb, es wird aber davon ausgegangen, dass dieses in naher Zukunft der Fall sein wird. Dann kann auch dafür garantiert werden, dass alle Dokumente des edoc-Servers zusätzlich in der DNB archiviert werden.

Für die digitalen Dissertationen wird ein derartiges Verfahren bereits seit einigen Jahren erfolgreich von der DNB durchgeführt. Die Dokumente sind dann noch einmal über den Deposit-Server in der DNB verfügbar.

## Fazit

Die Humboldt-Universität zu Berlin ist mit dem edoc-Server auf einem guten Weg, für digital publizierte Dokumente ein geeignetes Langzeitarchiv bereitzustellen. Obwohl das bisherige System (noch) nicht den Anforderungen der „Vertrauenswürdigen digitalen Langzeitarchive“, vgl. [8] entspricht und in jedem Detail dem OAIS-Modell [13] folgt, sind die Betreiber zuversichtlich, die Dokumente auch zukünftigen Generationen von Wissenschaftlern und Studierenden in einer benutzbaren Form anbieten zu können. Die weitere Arbeit an den technischen Details wird permanent fortgeführt. Im aktuellen Fokus steht dabei die Umsetzung einer geeigneten Metadatenstrategie, vgl. [11].

## Literatur

- [1] ROTHENBERG, JEFF: *Digital Information Lasts Forever – Or Five Years, Whichever Comes First*. 2. Oktober 2001, Vortrag, <http://www.amibusiness.com/dps/rothenberg-arma.pdf> (18.03.2009)
- [2] ARBEITSGRUPPE ELEKTRONISCHES PUBLIZIEREN: *Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin – Leitlinien* –, Humboldt-Universität zu Berlin. 2001, [http://edoc.hu-berlin.de/e\\_info/leitlinien.php](http://edoc.hu-berlin.de/e_info/leitlinien.php) (17.03.2009)
- [3] SCHWENS, UTE; LIEGMANN, HANS: *Die digitale Welt – eine ständige Herausforderung*. Kühlen, Rainer; Seeger, Thomas und Strauch, Dietmar, Handbuch zur Einführung in die Informationswissenschaft und -praxis, 5. völlig neu gefasste Ausgabe, Auflage, München, (2004)
- [2] ARBEITSGRUPPE ELEKTRONISCHES PUBLIZIEREN DINI : *DINI-Zertifikat für Dokumenten- und Publikationsservices 2007 [April 2007], DINI-Schriften 3, urn:nbn:de:kobv:11-10079197* (18.03.2009)
- [4] NESTOR - MATERIALIEN 8: *nestor – Kompetenznetzwerk Langzeitarchivierung / Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: nestor-Kriterien. Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 2, 2008, Frankfurt am Main : nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2008021802* (18.03.2009)
- [5] FROMM, NIELS: *Backup-Strategie für den Dokumentenserver, cms-journal 32, 2009, Humboldt-Universität zu Berlin, urn:nbn:XXXXXXXX* (16.03.2009)
- [6] REICH, VICKY; DAVID S. ROSENTHAL; LOCKSS: *A Permanent Web Publishing and Access System*. D-Lib Magazine, Volume 7 Number 6, June 2001, DOI: 10.1045/june2001-reich, <http://www.dlib.org/dlib/june01/reich/06reich.html> (20.02.2009)
- [7] PÉDAUQUE, ROGER T.: *Document: Form, Sign and Medium, As Reformulated for Electronic Documents*. (2003) URL: [http://archivesic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic\\_00000594.pdf](http://archivesic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic_00000594.pdf), (18.03.2009)
- [8] STANESCU, A.: *Assessing the Durability of Formats in a Digital Preservation Environment*. D-Lib Magazine (Band 10), Nr. 11., 2005, URL: <http://www.dlib.org/dlib/november04/stanescu/11stanescu.html> (17.03.2009)
- [9] PORTICO: <http://www.portico.org/> (17.03.2009)
- [10] FROMM, NIELS: *Signatur und Zeitstempel zur Wahrung von Authentizität und Integrität*. cms-journal 32, 2009, Humboldt-Universität zu Berlin, urn:nbn:XXXXXXXX, (16.03.2009)
- [11] FROMM, NIELS: *Einsatz elektronischer Signaturen auf dem Dokumentenserver der Humboldt-Universität zu Berlin*. Diplomarbeit, Humboldt-Universität zu Berlin, Institut für Informatik, 2009
- [12] SCHÖNING-WALTER, CHRISTA: *Der Uniform Resource Name (URN)*. In: nestor-Handbuch, Version 1.2 (Juni 2008), Kapitel 13.2.1, [http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_156.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_156.pdf) (18.03.2009)
- [13] CCSDS (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS): *Reference Model for an Open Archival Information System (OAIS)*. Blue Book ISO 14721: 2003 (Band Issue 1). URL: <http://www.ccsds.org/docu/dscgi/ds.py/Get/File-143/650xobi.pdf> (18.03.2009)

# Service für Autorinnen und Autoren

Maxi Kindling | maxi.kindling@ibi.hu-berlin.de

Peter Schirmbacher | schirmbacher@cms.hu-berlin.de

Die vorangehenden Kapitel haben einen Überblick zu den Entwicklungen und Möglichkeiten des wissenschaftlichen Publizierens im digitalen Zeitalter, verschiedenen Hilfsmitteln für das Erstellen eigener Arbeiten und dem Umgang des Publizierens nach den Kriterien von Open Access vermittelt. Die nachfolgenden Beiträge sollen konkrete Hilfestellung und Hinweise für das Publizieren auf dem Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin geben. In diesem Sinne werden die verschiedenen Aktivitäten und Angebote des edoc-Servers zur Unterstützung der Autorinnen und Autoren vorgestellt.

## Die Aktivitäten der Arbeitsgruppe Elektronisches Publizieren

Die Arbeitsgruppe Elektronisches Publizieren fand in diesem Heft bereits an verschiedenen Stellen Erwähnung. Sie setzt sich zusammen aus Mitarbeiterinnen und Mitarbeitern der Universitätsbibliothek und des Computer- und Medienservice sowie einer Reihe von Kolleginnen und Kollegen, die durch die Bearbeitung von Drittmittelprojekten zur Weiterentwicklung des Publikationsservice beitragen. Den Startschuss für die Arbeiten bildete 1997 ein Drittmittelprojekt zur elektronischen Veröffentlichung von Dissertationen. Auf der Basis des Datenbankmanagementsystems Sybase wurde eine Datenstruktur aufgebaut, die letztlich noch heute die Grundlage für die Mehrzahl der Aktivitäten darstellt. Von Beginn an war es wichtig, dass sich alle Entwicklungen nach internationalen Standards und Regelungen ausrichten.

Nicht selten wurden diese auch direkt durch die Arbeiten des Teams beeinflusst. Die Grundsätze des Herangehens sind in den Leitlinien [1] für den edoc-Server formuliert. Sie geben über die angestrebten Ziele einen sehr guten Überblick. Zeichen der Anerkennung der bisherigen Arbeiten ist die Platzierung des Servers seit mehr als zwei Jahren unter den Top 20 der weltweiten Repository-Rankingliste [2].

In gleicher Weise im Mittelpunkt der Bemühungen stand die Unterstützung der Autorinnen und Autoren auf ihrem Weg zu einer elektronischen Publikation. So existieren für jede der Veröffentlichungskategorien auf dem edoc-Server entsprechende Unterstützungstools, die die Arbeit erleichtern sollen.

Wie aus dem Screenshot (Abbildung 1) ersichtlich wird, teilt sich die Homepage des Dokumenten- und Publikationsservers in das Angebot für die Suchenden oder Lesenden auf der linken Seite und in das Angebot für die Autoren auf der rechten Seite, wobei hier unterschieden wird zum einen zwischen dem Angebot für Promovierende bzw. Habilitierende und zum anderen dem Angebot für so genannte Open-Access-Veröffentlichungen. Diese Tools sind auf den Webseiten des edoc-Servers sehr ausführlich beschrieben und sollen so zur unmittelbaren Nutzung einladen. Die nachfolgenden Artikel sind deshalb sehr knapp gehalten, weil sie im Wesentlichen auf das Webangebot hinweisen wollen.

*In den letzten 10 Jahren hat die Arbeitsgruppe EPUB eine Vielzahl von Aktivitäten zur Unterstützung der Autorinnen und Autoren auf ihrem Weg zum elektronischen Publizieren entwickelt. Der Beitrag verweist auf die wichtigsten Tools und Services und ist gleichzeitig als eine Klammer für die nachfolgenden Beiträge des cms-journals zu verstehen.*

*Arbeitsgruppe Elektronisches Publizieren, Service, Angebot, Dienstleistung, Autorenbetreuung*

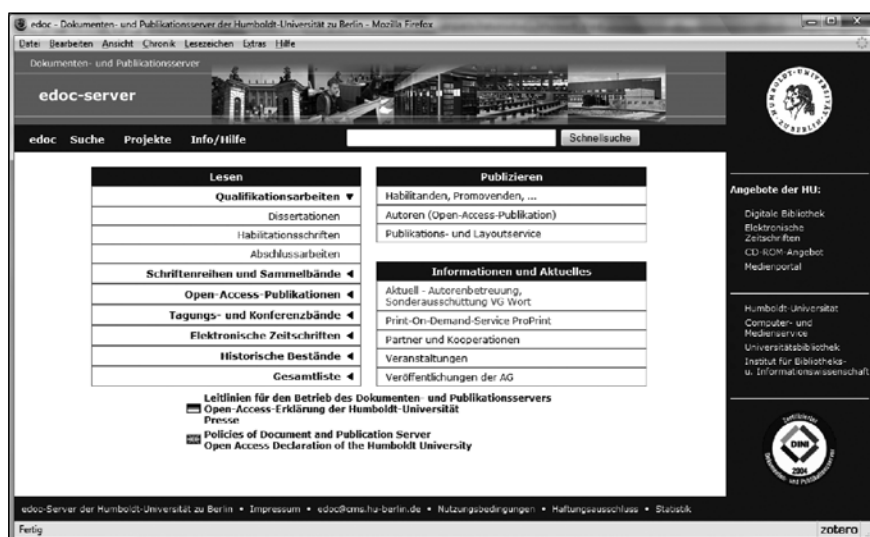


Abb. 1: Homepage des Dokumenten- und Publikationsservers der HU

## Die Benutzung von Dokumentvorlagen

Für nahezu alle Publikationsarten auf dem edoc-Server wurden und werden Dokumentvorlagen bereitgestellt als Basis für das Erstellen eigener Arbeiten. Sie sollen die Strukturierung des Textes erleichtern und sind von großer Bedeutung im Hinblick auf die Langzeitarchivierung der Arbeiten im Archivformat XML sowie für die Darstellung in einem Präsentationsformat wie html und pdf bzw. pdf/a. Mithilfe der Dokumentvorlage werden die Arbeiten nach einem standardisierten Verfahren strukturiert, indem z. B. Kapitel und Überschriften als solche gekennzeichnet werden. Die Dokumentvorlage beinhaltet dazu beispielsweise verschiedene Formatvorlagen und Makros. Bestes Beispiel ist die Dokumentvorlage für Dissertationen [3]. Ihre Nutzung wird nicht nur an der Humboldt-Universität als Bedingung für die Entgegennahme einer elektronischen Fassung der Dissertation vorausgesetzt, sondern sie wird an vielen Universitäten in Deutschland empfohlen. So ist sie auch der „Quasi“-standard innerhalb des durch die Deutschen Nationalbibliothek geführten Projekts „Dissertationen Online“. Im Beitrag [4] wird darauf näher eingegangen. Als Angebot für die Studierenden der Humboldt-Universität – aber natürlich auch darüber hinaus – bieten wir in dieser Rubrik weitere Dokumentvorlagen für Abschluss-

arbeiten wie Bachelor-, Master-, Diplom- oder Magisterarbeiten an. Es ist damit unser Ziel, möglichst vielen eine Handreichung zur Erleichterung des Publizierens zu bieten. Unsere Bemühungen gehen dahin, diese Dokumentvorlagen aktuell zu halten und sie den gerade verfügbaren Versionen der gängigen Textverarbeitungsprogramme anzupassen. Leider gelingt das aus Kapazitätsgründen nicht immer, sodass die Versionen für Word für Windows 2007 als auch für Open Office.org 3.0 noch nicht angeboten werden können. Wir arbeiten jedoch daran und sind optimistisch, dass sie im Verlauf des Sommers 2009 zur Verfügung stehen werden. Während die Bereitstellung der Formatvorlagen, verbunden mit der entsprechenden Beratung zu ihrer Benutzung, einen kostenfreien Service der Arbeitsgruppe darstellt, bieten wir ergänzend einen kostenpflichtigen Publikations- und Layoutservice für Dissertationen an, der in der letzten Zeit immer häufiger in Anspruch genommen wird.

Es ist nicht unser Ziel, Druck gegenüber den Autorinnen und Autoren zur Nutzung bestimmter Textverarbeitungssysteme auszuüben. Aus diesem Grund versuchen wir, uns auch auf die Gruppe der Latex-Nutzer einzustellen, die zwangsläufig in den naturwissenschaftlichen Instituten der Universität sehr verbreitet ist. In [5] werden dazu die entsprechenden Hinweise gegeben. Aufmerksam gemacht sei auch auf den in der Arbeits-

gruppe entstandene Kurs im Lernmanagementsystem Moodle (für alle mit einem HU-Moodle-Account siehe [6]).

## OA-Publikationen

Nach über 10 Jahren Existenz hat der edoc-Server das Anfängerstadium verlassen, sodass unser Angebot deutlich breiter geworden ist. Neben den Abschlussarbeiten findet man vermehrt elektronische Zeitschriften, herausgegeben von Wissenschaftlern und Wissenschaftlerinnen der Humboldt-Universität, Konferenzproceedings und Schriftenreihen einiger Institute der HU. Besonders hingewiesen sei jedoch auf die Rubrik Open-Access-Publikationen. Unter der Bezeichnung OA-Publikationen werden Pre- und Postprints zusammengefasst, die von den Wissenschaftlerinnen und Wissenschaftlern der HU auf dem edoc-Server zur Verfügung gestellt werden, nachdem sie an anderer Stelle, z. B. in einer wissenschaftlichen Zeitschrift, bereits erschienen sind oder das Erscheinen geplant ist. Natürlich ist diese Bezeichnung leicht irreführend, denn sämtliche Dokumente des Servers sind frei zugänglich. Trotzdem wollten wir diese besondere Kategorie herausstellen, weil sie ein Spiegelbild der wissenschaftlichen Arbeiten an der HU ist oder, korrekt formuliert, werden soll. Ähnlich dem Sammelauftrag der Universitätsbibliothek, der darin besteht, alle papierbasierten Veröffentlichungen von Angehörigen der HU zu sammeln, streben wir auf dem Server die Veröffentlichung der elektronisch verfügbaren Fassungen an. Wie man in drei Schritten zu seiner OA-Publikation kommt, lässt sich in [7] nachlesen.

### OA-Projektseminar zur Unterstützung der Autorinnen und Autoren

Im Sommersemester 2009 findet bereits zum dritten Mal ein Projektseminar „Publizieren nach den Kriterien von Open Access“ am Institut für Bibliotheks- und Informationswissenschaft statt. Ziel des Seminars ist es, zum einen interessierte Wissenschaftlerinnen und Wissenschaftlern bei der Klärung rechtlicher, organi-

satorischer und technischer Fragestellungen im Zusammenhang mit dem elektronischen Publizieren nach den Open-Access-Prinzipien behilflich zu sein und zum anderen den Studierenden die Möglichkeit zu geben, sich mit den unmittelbaren Problemen des Veröffentlichens im Detail auseinanderzusetzen. Jeder Studierende des Seminars „betreut“ eine oder mehrere Wissenschaftlerinnen bzw. Wissenschaftler auf dem Weg zur OA-Publikation. Das beginnt bei der Klärung rechtlicher Fragen im Zusammenhang mit einem möglichen Verlagsvertrag, setzt sich fort mit der Konvertierung des Artikels in ein „sauberes“ pdf/A-Format, was sich für diese Kategorie von Veröffentlichungen für die Langfristarchivierung eignet, und endet mit dem Upload des oder der Artikel auf dem edoc-Server, verbunden mit der Eingabe der notwendigen Metadaten.

Wenn es auch für eine Unterstützung bei der Veröffentlichung Ihrer Publikation auf dem edoc-Server in diesem Sommer zu spät sein mag, hoffen wir, Ihr Interesse für eines der kommenden Projekte geweckt zu haben. Weitere Informationen zum Ablauf des Seminars finden Sie auf den Webseiten des Instituts für Bibliotheks- und Informationswissenschaft unter: <http://www.ibi.hu-berlin.de/forschung/infomanagement/lehre/veranstaltungen/pseoa>. Über Ihre Rückmeldungen und Fragen freuen sich Uwe Müller ([u.mueller@cms.hu-berlin.de](mailto:u.mueller@cms.hu-berlin.de)) und Peter Schirmbacher ([schirmbacher@cms.hu-berlin.de](mailto:schirmbacher@cms.hu-berlin.de)).

## Literatur

- [1] [http://edoc.hu-berlin.de/e\\_info/leitlinien.php](http://edoc.hu-berlin.de/e_info/leitlinien.php)
- [2] [http://repositories.webometrics.info/top300\\_rep.asp](http://repositories.webometrics.info/top300_rep.asp)
- [3] [http://edoc.hu-berlin.de/e\\_autoren/index.php](http://edoc.hu-berlin.de/e_autoren/index.php)
- [4] HENNEBERGER, S.; SCHRÖDER, K.: *Elektronische Publikation von Dissertationen und Habilitationsschriften an der Humboldt-Universität zu Berlin*. cms-journal 32, Berlin 2009, S. 102–104
- [5] KLATT-KAFEMANN, M.; RYCKO, J.: *Eine gute Wahl: Wissenschaftliche Arbeiten mit LaTeX verfassen*. cms-journal 32, Berlin 2009, S. 106–108
- [6] <http://moodle.hu-berlin.de/course/view.php?id=2298>
- [7] *Hinweise zu Open-Access-Veröffentlichungen auf dem edoc-Server der Humboldt-Universität zu Berlin*. cms-journal 32, Berlin 2009, S. 105



Abb. 2: Layoutservice für MS Office – Staroffice/Openoffice – LaTeX



# Elektronische Publikation von Dissertationen und Habilitationsschriften an der Humboldt-Universität zu Berlin

Sabine Henneberger | shenneberger@cms.hu-berlin.de;

Karin Schröder | karin.schroeder@ub.hu-berlin.de

Durch Beschluss des Akademischen Senats ist an der Humboldt-Universität die Möglichkeit gegeben, der Veröffentlichungspflicht für Dissertationen auch durch die elektronische Publikation auf dem Dokumenten- und Publikationsserver der HU Berlin (edoc.hu-berlin.de) nach den Vorgaben der Universitätsbibliothek (UB) nachzukommen (Amtliches Mitteilungsblatt der HU Berlin Nr. 14/98).

Habilitationsschriften können freiwillig elektronisch publiziert werden. Die Publikation darf erst nach der erfolgreich abgelegten Disputation bzw. dem Rigorosum oder der Habilitation erfolgen.

Durch diesen Artikel bekommen Autoren wichtige Informationen zu den Themen:

- elektronische Version
- Abgabemöglichkeiten
- Druckexemplare
- weitere technische Hinweise
- Hilfsmöglichkeiten

## Die elektronische Version der Dissertation bzw. Habilitationsschrift

Die elektronische Version muss zwingend mit der Dokumentvorlage „dissertation-hu“ für Microsoft Word<sup>1</sup>, StarOffice oder OpenOffice geschrieben oder nachbearbeitet sein oder den Vorgaben und Styles für LaTeX-Dissertationen der Universitätsbibliothek entsprechen (weitere Informationen siehe unter LaTeXX).

*Für die elektronische Publikation von Qualifikationsarbeiten auf dem edoc-Server der Humboldt-Universität zu Berlin gelten bestimmte Bedingungen. Hinweise dazu gibt der folgende Artikel.*

*elektronisches Publizieren, edoc-Server der HU, Qualifikationsarbeit, Microsoft WORD, OpenOffice, StarOffice, Dissertation*

Die elektronische Version besteht aus:

- der Dissertation/Habilitationsschrift in einem der o. g. Originalformate inklusive Selbstständigkeitserklärung und digitalisierten Bildern und Grafiken, aber ohne Lebenslauf. Geben Sie immer das Datum der Einreichung der Dissertation/Habilitationsschrift beim Promotions-/Habilitationausschuss und das Datum der Promotion/Habilitation (d. h. Tag der mündlichen Prüfung bzw. Datum des Vortrags vor dem Fakultätsrat) sowie die Namen der Gutachter an!
- je einem Abstract in deutscher und englischer Sprache (max. 2000 Zeichen inkl. Leerzeichen, entspricht ca. ½ DIN-A4-Seite) und mindestens je vier Schlagwörtern in deutscher und englischer Sprache. Diese zusätzlichen Daten sind als einzelne Textdateien (Speicherung als txt-Datei) einzureichen, d. h. als Dateien, die als reine Textdateien ohne jede Formatierung (fett, kursiv, Exponenten, Indizes, Formeln, ...) und ohne Sonderzeichen (mit Ausnahme der direkt auf der PC-Tastatur enthaltenen) gespeichert sind.
- einem digitalen Druckexemplar im PDF-Format, das zusätzlich zu der digitalen Version im Originalformat abzugeben ist.

## Abgabe der elektronischen Version

- über die Abgabeseite im Internet, erreichbar über [http://edoc.hu-berlin.de/e\\_autoren/index.php](http://edoc.hu-berlin.de/e_autoren/index.php)
- nach Absprache Zusendung von CD-ROM (Postanschrift: Humboldt-Universität zu Berlin, Computer- und Me-

<sup>1</sup> Bei Microsoft Word kann auch die Dokumentvorlage „dissonline“, zu finden über [www.dissonline.de](http://www.dissonline.de) oder unter [http://edoc.hu-berlin.de/e\\_autoren/download/dissonline.dot](http://edoc.hu-berlin.de/e_autoren/download/dissonline.dot), verwendet werden.

dienservice, AG Elektronisches Publizieren, Rudower Chaussee 26, 12489 Berlin) möglich

- andere (z. B. E-Mail) nur in Ausnahmefällen und nur nach Absprache

#### Ansprechpartnerinnen für die elektronische Version

Arbeitsgruppe Elektronisches Publizieren:

Frau Schröder Tel. 2093-70074  
Frau Henneberger Tel. 2093-70075  
E-Mail: [edoc@cms.hu-berlin.de](mailto:edoc@cms.hu-berlin.de)

#### Papierexemplare der Dissertation für die Pflichtabgabe an die Deutsche Nationalbibliothek, Universitätsbibliothek und Sondersammelbibliothek (entfällt bei Habilitationsschriften)

- Es werden 4 gebundene Papierexemplare benötigt.
- Das verwendete Papier muss weiß, chlorfrei gebleicht und alterungsbeständig sein.
- Die Mindestanforderung an den Einband ist Schwarzpappenbroschur. Wenden Sie sich bitte an einen fachlich ausgewiesenen Buchbindebetrieb – **keinen Copyshop!**
- Folgende Bindungen können **nicht** als „dauerhaft haltbar“ **akzeptiert** werden: Ringbuchbindung, Kartonbroschur und sämtliche Bindemethoden, die unter Verwendung von Metall- oder Kunststoffteilen hergestellt wurden.
- Alternativ können die abzugebenden 4 Papierexemplare auch durch die UB hergestellt werden. Die Grundgebühr (Bindung) beträgt 43,45 Euro, die Vervielfältigungsgebühr 25 Cent pro geschriebener Seite (bzw. 6,25 Cent pro gedruckter Seite). Bei Arbeiten mit mehr als 1000 Seiten wird eine Zusatzgebühr von 30,65 Euro erhoben.

#### Ansprechpartner für die Druckexemplare

Hochschulschriftenstelle der Universitätsbibliothek: Tel. 2093-3221

## Technische Hinweise

Die Beschreibung der Aufbereitung und Formatierung Ihrer Arbeit und weitere Informationen finden Sie unter [http://edoc.hu-berlin.de/e\\_autoren/index.php](http://edoc.hu-berlin.de/e_autoren/index.php). Bei Fragen oder Problemen wenden Sie sich bitte an die AG Elektronisches Publizieren. Die folgenden Ausführungen stellen lediglich allgemeine Hinweise dar.

#### Für Microsoft Word, StarOffice und OpenOffice

Wenn Sie Ihre Dissertation/Habilitationsschrift mit den Textverarbeitungssystemen Microsoft Word für Windows (oder MacOS), StarOffice oder OpenOffice schreiben, ist es notwendig, die Dokumentvorlage „dissertation-hu“ zu benutzen. Beim Schreiben mit Microsoft Word ist auch die Verwendung von „dissonline.dot“ möglich. Mit Hilfe dieser Dokumentvorlage nehmen Sie eine strukturierte Formatierung der Dissertations-/Habilitationsschrift vor, indem Sie den einzelnen Teilen Ihrer Arbeit die dafür vorgesehenen Formatvorlagen zuweisen. Diese Strukturierung ermöglicht es uns, Ihre Arbeit in einer Datenbank über das Internet zur Verfügung zu stellen und recherchierbar zu machen sowie die Langzeitarchivierung zu gewährleisten.

Wir können keine Arbeiten akzeptieren, die nicht nach diesen Vorgaben aufbereitet sind. Informieren Sie sich möglichst schon vor Beginn des Schreibens!

Die Dokumentvorlage kann über die Adresse [http://edoc.hu-berlin.de/e\\_autoren/vorlage.php](http://edoc.hu-berlin.de/e_autoren/vorlage.php) heruntergeladen werden. Dort erhalten Sie auch weitere Informationen, wie Hinweise zur Installation der Dokumentvorlage, eine Beschreibung und eine Checkliste zur Erfüllung unserer Vorgaben.

Für bereits geschriebene Qualifikationsarbeiten ist eine Nachformatierung mit der Dokumentvorlage möglich. Es kann aber u. U. einen größeren Aufwand erfordern. Das betrifft insbesondere das Literaturverzeichnis. Planen Sie dafür bitte ausreichend Zeit ein.

Über die Berufliche Weiterbildung der Humboldt-Universität werden unter der Bezeichnung „CMS-DSW Dissertationen schreiben mit Microsoft Word (elektronische Veröffentlichung)“ Kurse für den Umgang mit der Dokumentvorlage [dissertation-hu.dot](http://dissertation-hu.dot) bzw. [dissonline.dot](http://dissonline.dot) angeboten.

Wenn Sie Schwierigkeiten bei der Nachformatierung Ihrer Dissertation haben, können wir Ihnen alternativ den kostenpflichtigen Formatierungs- und Druckservice der Humboldt-Innovation GmbH anbieten ([http://edoc.hu-berlin.de/formatierungsservice/e\\_autoren/formatierung.php](http://edoc.hu-berlin.de/formatierungsservice/e_autoren/formatierung.php)). Die Mitarbeiterinnen übernehmen für Sie die notwendigen Formatierungen Ihrer elektronischen Publikation. Sie liefern Ihre Worddatei, Gliederungsanweisungen, alle Grafiken und Tabellen.

#### Für LaTeX

Bei der Abgabe von LaTeX-Dateien ist die Nutzung der speziell dafür entwickelten Vorlage notwendig. Es sind die Richtlinien zur Erstellung von LaTeX-Arbeiten einzuhalten, die im Leitfaden unter [http://edoc.hu-berlin.de/e\\_autoren/latex](http://edoc.hu-berlin.de/e_autoren/latex) zu finden sind. Zur Erzeugung der PDF-Version wird die Verwendung von pdf<sub>l</sub>atex vorausgesetzt. Es ist darauf zu achten, dass alle zur Kompilation notwendigen Dateien (vor allem die Quell-Dateien – TEX und BIB – und Bilder) mit abgegeben werden. Verwendete Grafiken sind jeweils als EPS- und PDF-Dateien abzugeben. Die für das Literaturverzeichnis verwendete Stil- und DVI-Dateien müssen mit abgegeben werden. Vor der Abgabe sollten Sie Ihre Dateien anhand der Checkliste überprüfen, die im Leitfaden enthalten ist. Bei auftretenden Fragen wenden Sie sich bitte an die LaTeX-Betreuung per E-Mail [edoc-latex@cms.hu-berlin.de](mailto:edoc-latex@cms.hu-berlin.de)

#### Andere Textverarbeitungssysteme

Andere Textverarbeitungssysteme unterstützen wir nicht.

## Einbinden von Multimediaelementen

Sie haben die Möglichkeit, Audio- und Videosequenzen oder andere aktive Komponenten einzubinden. Bitte wenden Sie sich an uns, da hier spezielle Regelungen getroffen werden müssen.

## Erstellen der PDF-Datei

Zur Erstellung der PDF-Version wird die Software „Adobe Acrobat“, mindestens in der Version 5.0, benötigt. Dieses Programm bietet umfangreiche Einstellungsmöglichkeiten bezogen auf die technische Qualität. Die für die elektronische Veröffentlichung bestimmte PDF-Datei muss Kriterien für eine Verarbeitung durch Print-on-Demand-Dienstleister sowie zur Archivierbarkeit erfüllen. Schreiben Sie mit Microsoft Word 2007, können Sie das Tool „Save as PDF“ der Firma Microsoft benutzen. Ausführliche Hinweise erhalten Sie unter [http://edoc.hu-berlin.de/e\\_autoren/download/Erstellung.pdf](http://edoc.hu-berlin.de/e_autoren/download/Erstellung.pdf)

Weitere Informationen finden Sie unter der Internet-Adresse des Dokumenten- und Publikations-servers:

<http://edoc.hu-berlin.de>

# Lehrgangsübersicht des CMS 2009

Das gemeinsame Weiterbildungsangebot der Beruflichen Weiterbildung und des Computer- und Medienservice finden Sie unter BWb-online (<http://www2.hu-berlin.de/berweit/BWb/>). Das aktuelle Kursangebot des CMS entnehmen Sie bitte der folgenden Aufstellung. Weitere Informationen sind unter der angegebenen Webadresse einsehbar.

## Anwendungssoftware

CMS-NZC	Nutzung des zentralen Computeservice im CMS
CMS-NZA	Nutzung des globalen Andrew Filesystems (AFS)
CMS-IAF	Installation einer Andrew Filesystem-Zelle im Institut
CMS-SPS	Einführung in SPSS
CMS-DSW	Digitale Dissertationen schreiben mit Microsoft WORD
CMS-TYG	Typografisches Gestalten
CMS-PCS	Photoshop-Praxis: Alles rund um Pixel, Auflösung, und Bildformate

## Multimedia Lehr- und Lernzentrum

CMS-DTL	Einsatz digitaler Technologien in der Lehre
CMS-OKK	Online kommunizieren und kooperieren in Lehre und Forschung
CMS-OTA	Onlinetests und Assessments: Wissen online überprüfen
CMS-PLA	Präsenzveranstaltungen mit e-Learning-Angeboten anreichern

## WWW

CMS-PUF	Forum für Plone-User
CMS-PKU	Plone-Kurs für Version 2.5

## Medientechnik

CMS-MHT	Einführung in die Medien- und Hörsaaltechnik des Erwin Schrödinger-Zentrums
---------	---

## Videotechnik und -bearbeitung

CMS-EVT	Einführung in die Videotechnik
CMS-DVS	Digitaler Videoschnitt mit Final Cut Pro

Zusätzlich wird angeboten, Kurse nach gesonderter Vereinbarung bei Ihnen im Institut zu halten. Ansprechpartner sind Frau Schloßecki, Tel.: 2093-70012, E-Mail: [jutta.schlossecki@cms.hu-berlin.de](mailto:jutta.schlossecki@cms.hu-berlin.de) und Herr Stange, Tel.: 2093-70040, E-Mail: [stange@cms.hu-berlin.de](mailto:stange@cms.hu-berlin.de).

# Hinweise zu Open-Access-Veröffentlichungen auf dem edoc-Server der Humboldt-Universität zu Berlin

## Klärung der rechtlichen Rahmenbedingungen

Mit der Erfassung auf dem Dokumenten- und Publikationsserver verbleiben die Urheberrechte bei Ihnen. Die Einhaltung von Urheber- und Verwertungsrechten Dritter liegt in der Verantwortung der Autoren bzw. der Herausgeber der elektronischen Dokumente.

Die AG Elektronisches Publizieren benötigt von Ihnen die Einwilligung, dass Ihre Publikation über den edoc-Server im Internet zur Verfügung gestellt wird.

Dabei muss insbesondere ermittelt werden, ob und unter welchen Bedingungen eine Zweitveröffentlichung zulässig ist. Dies hängt in hohem Maße von den jeweils getroffenen Vereinbarungen zwischen Verlagen und Autoren ab. Die unterschiedlichen Regelungen reichen von einem generellen Verbot der Weiterverbreitung des publizierten Artikels – in diesem Fall ist keine Open-Access-Publikation möglich – über eine zeitliche Befristung dieser Ausschließlichkeit bis hin zur Erlaubnis, die Artikel auf einem Webserver zu veröffentlichen, der zur Institution des Autors gehört.

Bitte holen Sie sich auch das Einverständnis evtl. vorhandener Koautoren ein.

Hinweise zu Veröffentlichungsrechten finden Sie auf folgenden Seiten im Internet:

<http://www.dini.de/oap/>

<http://www.sherpa.ac.uk/romeo.php?all=yes>

## Technische Hinweise

Für die Veröffentlichung auf dem edoc-Server ist eine PDF-Datei, die den eigentlichen Artikel enthält, erforderlich. Diese darf keinen Sicherheitsbeschränkungen unterliegen und damit auch nicht durch ein Kennwort geschützt sein. Es ist der PDF/A-Standard einzuhalten, welcher besagt, dass alle zugehörigen Bestandteile in der Datei enthalten sind. Das heißt für Sie, dass verwendete Schriften (Fonts) eingebettet sein müssen und Grafik-, Video- und Audiodateien nicht als externe Dateien vorliegen, sondern ebenfalls in der Datei enthalten sind.

Weiterhin werden von Ihnen Schlagwörter in Deutsch und Englisch und Angaben, wann und wo original publiziert wurde, benötigt.

Falls ein Abstract existiert, sollte dieser auch zur Verfügung gestellt werden.

Falls vorhanden, kann eine Textverarbeitungsdatei oder Textsatzdatei (z. B. Microsoft Word, LaTeX) vorgelegt werden, die keinen Sicherheitsbeschränkungen unterliegt, nicht durch ein Kennwort geschützt ist, alle zum Inhalt gehörigen Teile (wie Grafik-, Video-, Audiodateien) enthält und nach Absprache mit Hilfe einer Dokumentvorlage oder entsprechend den Stylevorgaben der Arbeitsgruppe Elektronisches Publizieren am Computer- und Medienservice der Humboldt-Universität formatiert wurde.

## Abgabeformalitäten

Die Abgabe der Publikation erfolgt, indem auf dem edoc-Server das Formular zum Übertragen der Dateien benutzt wird. Sie finden es unter:

<https://edoc.hu-berlin.de/MetaIn/oa/index.php>

Neben der Datei, die das eigentliche Werk enthält, sind an dieser Stelle einige zusätzliche Informationen (Metadaten) zur Beschreibung des Artikels einzugeben: die Namen der Autoren, der Titel und das ursprüngliche Erscheinungsdatum der Publikation, Angaben zur Zeitschrift (Name, Heftnummer, Seitenangabe) und eine grobe fachliche Einordnung. Schlagwörter und ein Abstract können hinzugefügt werden.

Wenn Sie Publikationen über diese Schnittstelle übertragen, stimmen Sie deren Bereitstellung im Sinne der Open-Access-Erklärung der Humboldt-Universität auf dem edoc-Server zu und bestätigen, dass alle beteiligten Autoren damit einverstanden sind.

Bei Fragen zur Open-Access-Veröffentlichung hilft Ihnen die AG Elektronisches Publizieren gerne weiter, die Mail-Adresse lautet:

[openaccess@hu-berlin.de](mailto:openaccess@hu-berlin.de)

Weitere Informationen finden Sie auch auf den Seiten der AG Elektronisches Publizieren im Internet:

[http://edoc.hu-berlin.de/e\\_autoren/index-oa.php](http://edoc.hu-berlin.de/e_autoren/index-oa.php)

# Eine gute Wahl: Wissenschaftliche Arbeiten mit LaTeX verfassen

Joanna Ryćko | joanna.rycko@cms.hu-berlin.de

Manuel Klatt-Kafemann | manuel.klatt@cms.hu-berlin.de

## Arbeitsweise vs. Arbeitsumgebung

Eine grundlegende Fragestellung für das Verfassen wissenschaftlicher Arbeiten ist die Wahl des richtigen Werkzeugs – unabhängig davon, ob es um das Schreiben einer kurzen Seminararbeit geht oder die möglicherweise monatelange Arbeit an einer Dissertation. Jeder Mensch hat diesbezüglich eine andere Herangehensweise und es steht verschiedene Software zur Verfügung, die die Bedürfnisse des Einzelnen erfüllt.

Wie sollte beim Erstellen einer wissenschaftlichen Arbeit vorgegangen werden? Am Anfang steht zumeist das leere Blatt Papier. Manche Verfasser<sup>1</sup> beginnen allgemein und verfeinern den Text nach und nach. Andere gehen eher chaotisch vor und schreiben etwas Text oder nur Stichworte; sortieren diese Aufzeichnungen, schreiben sie um und ergänzen. Einige fertigen zu Beginn eine Gliederung an und versuchen, auf dieser aufzubauen. Nicht selten kommt es vor, dass der Text im Verlauf des Schreibens in ganz andere Richtung geht, als zu Beginn gedacht. Dies liegt darin begründet, dass der Text letztlich die Recherche, die der Verfasser durchführt, widerspiegelt, sie ordnet, zusammenfasst und darstellt. Mitunter ist die Recherche abgeschlossen und die Ergebnisse sind bereits präsentiert, bevor überhaupt mit dem Schreiben begonnen wird (so passiert es zum Beispiel bei Ausarbeitungen in Seminaren); manchmal geschieht dies parallel. Auch die Literaturquellen als Ergebnis der Recherche werden in unterschiedlicher

Weise gesammelt bzw. erfasst. Geht man dabei nicht systematisch vor, liegen schließlich dutzende Einträge z. T. unvollständig und ungeordnet vor und es lässt sich kaum ein brauchbares Literaturverzeichnis erstellen. Für die Wahl einer Software zur Unterstützung des Schreibens ist daher in jedem Fall zu beachten, dass sich der Verfasser ganz auf das Schreiben und den Inhalt der Arbeit konzentrieren kann und die Arbeitsweise des Programms die eigene Arbeitsweise unterstützt.

Es gibt keine perfekte Software für „Jedermann“, soviel sei an dieser Stelle zusammengefasst. Im Folgenden soll daher ein Werkzeug vorgestellt werden, das der Arbeitsweise der Autoren dieses Artikels und möglicherweise auch der des Lesers entspricht.

Von ebenso großer Bedeutung wie die Arbeitsweise ist die Auseinandersetzung mit den Merkmalen und Funktionalitäten der einzelnen Programme und es lohnt sich sicher, dieser Aufgabe etwas Zeit zu widmen.

Die hier vorgestellte Arbeitsumgebung basiert auf LaTeX – „basiert“, weil LaTeX nur einen Bestandteil darstellt. LaTeX ist als untere Schicht, quasi als „Engine“ zu sehen. Die darauf aufsetzende Schicht bilden ein beliebiger Texteditor und ein Vorschau-Programm. Sowohl Texteditoren, entsprechende Vorschau-Programme als auch TeX-Distributionen gibt es für jedes Betriebssystem, d. h. eine solche Arbeitsumgebung ist plattformunabhängig. Empfehlungen finden Sie am Ende dieses Artikels.

*In diesem Artikel wird das Textverarbeitungsprogramm LaTeX im Hinblick auf das Verfassen wissenschaftlicher Arbeiten besprochen. Dabei werden die Vorteile von LaTeX im Vergleich zu klassischen Textverarbeitungsprogrammen aufgezeigt sowie Software-Empfehlungen gegeben.*

*LaTeX, Dissertation, Dokumentvorlage*

<sup>1</sup> Es wird allein aus Gründen der besseren Lesbarkeit die maskuline Form verwendet.



## Was ist LaTeX?

LaTeX ist ein Textsatzsystem und ein Teil eines größeren Systems namens TeX.<sup>2</sup> Der grundlegende Unterschied zwischen LaTeX als Textsatzsystem und klassischen Textverarbeitungsprogrammen wie z. B. Open Office Writer und Microsoft Word besteht darin, dass beim Schreiben die Formatierung des Textes in LaTeX nicht sichtbar ist, sondern nur die reine Textdatei. Es handelt sich nicht wie z. B. bei Word um ein WYSIWYG<sup>3</sup>-System, wo sämtliche Formatierung, Grafiken und andere Auszeichnungen am Bildschirm direkt beim Schreiben betrachtet werden können. Das Prinzip ähnelt eher (X)HTML. In LaTeX werden Teile des Textes, die man abheben möchte (z. B. als Überschrift oder kursiv setzen) mit speziellen Befehlen gekennzeichnet. Diese Befehle sind während des Schreibens sichtbar und es lassen sich auch ihre Grenzen erkennen. Beispiel:

```
\chapter{Titel des neuen Kapitels}
```

– mit diesem Befehl wird die Überschrift eines Kapitels gesetzt. Die Klammern zeigen, wo der Titel beginnt und endet. Wie das Ganze im endgültigen Dokument aussehen wird, ist beim Schreiben zunächst nebensächlich, da dies jederzeit angepasst werden kann und somit die Konzentration vorerst allein dem Inhalt gilt. LaTeX ermöglicht zudem, sich eigene Befehle auszudenken und diese zu definieren. Somit lassen sich zum Beispiel Wörter von besonderer Bedeutung markieren und ihre Darstellung später anpassen.

## Arbeiten mit großen Dateien

Bei wissenschaftlichen Arbeiten haben wir es meist mit langen Dokumenten zu tun, die besondere Herausforderungen an die Arbeitsumgebung darstellen – vor allem, dann, wenn im Dokument mehrere Bilder vorkommen. Je aufwendiger das Dokument, desto langsamer wird es sich öffnen und seine Handhabung wird

mühsam. Auch das Auffinden einer bestimmten Stelle im Dokument kann dann schwerfallen. Bei klassischen Textverarbeitungsprogrammen kann dies zu sehr tragem Arbeiten bis hin zu Abstürzen des Arbeitssystems führen. Je nach System werden dabei proprietäre Dateiformate verwendet, die kein Aufsplitten ermöglichen und die zum Teil beim Absturz unwiederbringlich beschädigt werden. In LaTeX arbeitet man mit Textdateien. Das bedeutet u. a., dass es keine Möglichkeit gibt, die Bilder direkt im Texteditor anzuzeigen, sondern es werden im Text bestimmte Stellen markiert, wo diese Grafiken zu platzieren sind. Die Größe der Bilder hat demnach keinen Einfluss auf das Öffnen und Benutzen des Dokumentes. Auch wenn eine LaTeX-Quelldatei abstürzen sollte, ist sie weiterhin lesbar und ein Teil oder die ganze Datei lässt sich retten.

Eine gern benutzte Funktion ist die der Splittung des Dokuments in mehrere kleinere Teile (z. B. in Kapitel) und das Aufrufen dieser von einer Hauptdatei. Es gibt Editoren (z. B. TeXnicCenter), die diese Funktion unterstützen und die Dateien in einer Baumdarstellung anzeigen. Dabei besteht die Möglichkeit, in der Hauptdatei festzulegen, an welchem Teil gerade gearbeitet wird. Alle anderen werden dann vorübergehend „ausgeschaltet“. Dank dessen lässt sich schnell und effizient arbeiten – ohne die Angst vor einem Absturz.

## Verzeichnisse

Am Anfang dieses Artikels wurde bereits das Literaturverzeichnis erwähnt. Zu den LaTeX-Tools gehört u. a. auch das Programm BibTeX, mit dem man Literaturverzeichnisse erstellen kann. Die Literaturdatenbank ist dabei immer eine Textdatei, die in besonderer Weise strukturiert ist. Diese Struktur kann man entweder per Hand editieren oder von einem Programm verwalten lassen, was eine bequeme Lösung ist. Je nach Tool sind aber nicht alle BibTeX-Optionen verfügbar. In jedem Fall ist es empfehlenswert, sich bereits zu Beginn des Schreibens der wissenschaftlichen Arbeit eine solche Literaturdatenbank anzulegen und diese systematisch zu ergänzen.

In LaTeX werden viele Verzeichnisse automatisch erstellt, es reicht meistens, einen Befehl an der gewünschten Stelle einzufügen. Möglich sind z. B. Inhaltsverzeichnis, Tabellen-, Bilderverzeichnis, Literaturverzeichnis, Stichwortverzeichnis usw. – alle diese lassen sich in LaTeX automatisch erstellen. Um das Layout muss man sich meist nicht kümmern, da es für fast jeden Anwendungsfall und jede Fachrichtung entsprechend vordefinierte Stile nach typografischen und z. B. beim Literaturverzeichnis nach den gängigen fachspezifischen Zitierregeln gibt. Für eine Anpassung der Stile an den eigenen Geschmack oder besondere Bedürfnisse lassen sich so genannte Pakete einfügen und deren Funktionen benutzen.

## Mathematik

LaTeX ist bekannt für seine Möglichkeit, komplizierte mathematische und chemische Formeln „schön“ und typografisch korrekt darzustellen. Das Eintippen dieser Formeln ist dabei relativ einfach und logisch. Nach einer kurzen Einarbeitungszeit lassen sich Formeln schnell mithilfe der Tastatur schreiben – es bleiben im Verhältnis zu anderen Systemen viele Mausklicks erspart bzw. es ist überhaupt erst möglich, diese Formeln zu schreiben. Aus diesem Grund greifen offenbar besonders viele Naturwissenschaftler zu LaTeX.

## Editieren und Vertreiben

Bei üblichen WYSIWYG-Textverarbeitungsprogrammen wird meist ein und dieselbe Datei sowohl für die Eingabe als auch für die Ausgabe verwendet. Die zwei Etappen Editieren und Vertreiben (Publizieren) können also im selben Format erfolgen. Dies hat Vorteile, bringt aber auch Nachteile, wenn auf einem anderen Rechner die ursprünglich verwendeten Schriften oder Drucker nicht installiert sind und somit das Layout und Erscheinungsbild beim Betrachter sich ändert – bis hin zur Ersetzung durch andere Schriftarten, die z. B. bei chinesischen Schriftzeichen zu einem

2 Entsprechende Hinweise finden sich am Ende des Beitrags.

3 WYSIWYG: What You See Is What You Get – was Sie am Bildschirm sehen, ist, was Sie ausgedruckt erhalten

völlig veränderten Inhalt führen können. Eine Lösung dieses Problems wäre der Vertrieb in einem anderen Format, z. B. PDF. Genau dies wird in LaTeX praktiziert, die beiden Etappen sind voneinander getrennt und werden durch unterschiedliche Dateiformate realisiert: Zum Editieren reicht eine Textdatei aus. Diese kann mit jedem Editor unter jedem gängigem Betriebssystem geöffnet werden. Einzig das Einstellen der ursprünglichen Zeichenkodierung (die je nach Editor automatisch und korrekt erkannt wird) ist zu beachten. Zum Vertreiben des Dokumentes wird die PDF-Datei verwendet, die aus der Textdatei direkt erstellt werden kann. Dies geschieht mittels des in jeder TeX-Distribution enthaltenen Programms `pdflatex`. Je nach Einstellung kann diese PDF-Datei z. B. für den Druck oder für die Langzeitarchivierung (nach dem Standard PDF/A-1a) optimiert werden, was bei wissenschaftlichen Arbeiten eine bedeutende Rolle spielt, da das niedergeschriebene Wissen langfristig verfügbar sein soll. Die einfache Textform der Dokumentquellen ermöglicht es, das Dokument nach Bedarf kooperativ zu erstellen (oder auch die Korrekturen der Gutachter direkt in den Quellcode einzutippen). Diese kann nachher mittels eines Versionsierungssystems (z. B. Subversion) und eines speziellen Programms (z. B. diff) mit der originalen Version verglichen werden. Die erneute Erzeugung der PDF-Datei auf einem anderen Rechner setzt jedoch auch hier das Vorliegen aller verwendeten LaTeX-Pakete und Schriften voraus. Unter der Voraussetzung, dass Standardpakete und Schriften verwendet wurden, lassen sich jedoch 25 Jahre alte Dokumente auch heute noch problemlos in ein Präsentationsformat wie PDF umwandeln.

## Freie Software

Nicht ohne Bedeutung ist auch die Tatsache, dass sowohl das Programm LaTeX als auch alle anderen Tools, die benötigt werden, um eine Arbeit mit LaTeX zu erstellen, als freie Software zugänglich sind. Das bedeutet unter anderem, dass sie kostenlos zur Verfügung stehen. Es gibt eine Vielzahl von kostenlosen Edi-

toren, die für LaTeX geeignet sind und dies besonders gut unterstützen, z. B. TeXnicCenter.

## Kein Vorteil ohne Nachteil: Tabellen

Bei allen Vorteilen hat LaTeX auch Nachteile, die manchen von der Benutzung abschrecken. Das größte Problem bereitet meist das Erstellen von komplexen Tabellen. Diese müssen – wie alle anderen Quellcode-Teile – auch in der Textdatei eingegeben werden. Auch wenn man dabei auf die Formatierung des Textes achtet (z. B. durch das Einrücken der Zeilen, das Trennen einzelner Zellen mit Leerzeichen usw.), wird es bald unübersichtlich. Einige Editoren helfen dabei, Tabellen mit LaTeX zu erstellen, indem sie spezielle Oberflächen dazu bereitstellen. Allerdings ist es meistens nur für einfache Tabellen anwendbar und das auch nur beim ersten Erstellen einer Tabelle. Das Editieren erfolgt anschließend via Quellcode.

## Empfehlungen

Falls Ihr Interesse geweckt ist, LaTeX auszuprobieren, empfehlen sich für den Einstieg folgende Kombinationen:

- unter Windows: als Distribution nehmen Sie MikTeX oder TeX Live 2008; als Editor eignet sich am besten TeXnicCenter, den man in wenigen Schritten für die Zusammenarbeit mit dem zuvor installierten LaTeX konfigurieren kann;
- unter Linux: Sie können ebenfalls TeX Live 2008 nehmen oder auch die mit Ihrem Linux-Paketen gelieferte TeX-Version; als Editor können Sie zum Beispiel kile ausprobieren;
- unter Mac OS X: hier gibt es ebenfalls TeX Live 2008 oder auch MacTeX, das sich bequem installieren lässt, der am meisten verbreitete Editor unter Mac ist TeXShop, mit dem man zwischen den Quellen und dem Ergebnis (PDF-Datei) hin und her springen kann.

Um mehr über LaTeX zu erfahren, können Sie zu zahlreichen Literaturquellen greifen oder aus reichlich vorhandenen Skripten auswählen, die im Internet verfügbar sind. Eine erste Anlaufstelle sind die auf dem edoc-Server angebotenen Informationen [1] inklusive einer ausgefeilten, insbesondere für wissenschaftliche Arbeiten angedachten Formatvorlage. Diese Vorlage unterstützt die Autoren dabei insbesondere durch eine vorbereitete Aufteilung in Kapitel, Deckblatt, Anhänge und Vorwort sowie auch bei der korrekten Optionssetzung für die PDF-Erzeugung inklusive der Dokumenteigenschaften. Gängige Literaturstile werden unterstützt und auch ein Beispiel für die korrekte Verwendung ist angegeben. Zusätzlich kann der im Moodle-System [2] angebotene Kurs als Einstieg in LaTeX dienen. Eine weitere gute Anlaufstelle ist CTAN<sup>4</sup>, ein Server, auf dem unterschiedliche LaTeX- (und TeX-) bezogene Dateien (Pakete, Dokumentationen, Programme) gespeichert sind und zur Verfügung gestellt werden. Es gibt weltweit mehrere Mirrors von CTAN, die einen schnelleren Zugriff auf die Inhalte ermöglichen. Innerhalb von Deutschland empfiehlt es sich z. B., die Server des deutschsprachigen Vereins der TeX-Anwender „Dante“<sup>5</sup> zu nutzen.

Bei Fragen können Sie sich an eine der zahlreichen Diskussionsgruppen und Mailinglisten wenden bzw. steht Ihnen das LaTeX-Betreuungsteam als Ansprechpartner gern zur Verfügung.

Kontakt: Tel. 2093-70003,  
E-Mail: [edoc-latex@cms.hu-berlin.de](mailto:edoc-latex@cms.hu-berlin.de)

## Literatur

- [1] [https://edoc.hu-berlin.de/e\\_autoren/latex/](https://edoc.hu-berlin.de/e_autoren/latex/)
- [2] <https://lms.hu-berlin.de/moodle/course/view.php?id=2298>

<sup>4</sup> URL: <http://www.ctan.org/>

<sup>5</sup> URL: <http://www.dante.de/>

# Literaturverwaltung zur Unterstützung des wissenschaftlichen Publizierens

Maxi Kindling | maxi.kindling@ibi.hu-berlin.de

Matti Stöhr | matti.stoehr@ibi.hu-berlin.de

## Literaturverwaltung und Literaturverwaltungssysteme

Literaturverwaltung im wissenschaftlichen Publikationsprozess meint die Organisation der für eine Publikation und/oder ein Arbeitsprojekt verwendeten Literatur. Sie umfasst das Anlegen eines Datenpools aus bibliografischen Angaben und/oder Literatur sowie deren formale und inhaltliche Erschließung; die Erstellung von Bibliografien bzw. Literaturverzeichnissen zu einer Veröffentlichung oder einem Projekt sowie die Organisation von Zitaten, Referenzen, Anmerkungen, Notizen und Gedanken.

Während Umberto Eco in seinem Standardwerk zum wissenschaftlichen Schreiben von 1977 für den Umgang mit Publikationen im Erstellungsprozess einer wissenschaftlichen Arbeit noch die Verwendung von so genannten Lesekarteien bzw. Lektüre-Karten empfiehlt<sup>1</sup>, hat sich die Literaturverwaltung den elektronischen Arbeitsumgebungen angepasst und es steht heute eine Vielzahl verschiedener Werkzeuge zur Verfügung.<sup>2</sup>

Ein (rechnergestütztes) Literaturverwaltungssystem bzw. eine Literaturverwaltungssoftware dient in erster Linie dem erleichterten Umgang nicht nur mit Literaturquellen, sondern auch der Strukturi-

rierung der eigenen literaturgestützten Arbeit – unabhängig von deren Zielstellung bzw. Ergebnis.

## Anwendungsformen von Literaturverwaltungssystemen

Zunächst werden so genannte „lokale“ Literaturverwaltungssysteme von webbasierten Literaturverwaltungssystemen unterschieden. Erstere werden auf dem eigenen Rechner installiert (Einzelplatz- bzw. Stand-Alone-Anwendung) und sind als „geschlossene“ Systeme mit einem beschränkten Nutzerkreis (in der Regel eine Person) konzipiert.<sup>3</sup> Der Zugriff auf webbasierte Literaturverwaltungssysteme erfolgt mithilfe eines Internetbrowsers. Sie speichern die abgelegten Daten nicht auf dem eigenen Rechner, sondern sie werden auf einem externen Webserver abgelegt. Webbasierte Literaturverwaltungssysteme<sup>4</sup> haben den Vorteil, dass man ortsunabhängig von einem beliebigen, internetfähigen Rechner auf das System zugreifen kann und ggf. mehrere Personen kollaborativ an einer Literatursammlung arbeiten können. Hierbei sind zwei Formen der Zusammenarbeit zu nennen: Zum einen kann der Datenpool für mehrere Benutzer freigegeben werden und zum anderen können Nutzerzugänge „geteilt“ werden. Mithilfe der Freigabe eines Datenpools für mehrere Benutzer bzw. einer Anwendung, mit der die gespeicherten Publikationen samt ihrer Me-

*Die gemeinsame Arbeitsgruppe Elektronisches Publizieren von UB und CMS hat sich mit dem grundsätzlichen Einsatz von Literaturverwaltungssystemen im wissenschaftlichen Publikationsprozess auseinandergesetzt und dabei einzelne Anwendungen bzw. Anwendungsformen untersucht. Der folgende Beitrag bietet einen kurzen Überblick zu bestehenden Formen von Literaturverwaltungssystemen und ihren Funktionen.*

*Literaturverwaltung, Literaturverwaltungssysteme, Wissenschaftliches Arbeiten*

1 vgl. Eco, Umberto (2007) Wie man eine wissenschaftliche Abschlussarbeit schreibt. Stuttgart: UTB (12. Auflage, Originalausgabe von 1977).

2 Eine sehr gute Übersicht zu Literaturverwaltungsprogrammen stellen Thomas Stöber und Astrid Teichert (Universitätsbibliothek Augsburg) bereit: <http://www.bibliothek.uni-augsburg.de/service/literaturverwaltung/> bzw. es lohnt sich auch ein Blick in die englischsprachige Wikipedia: [http://en.wikipedia.org/wiki/Comparison\\_of\\_reference\\_management\\_software](http://en.wikipedia.org/wiki/Comparison_of_reference_management_software) [Der letzte Zugriff auf alle im Artikel referenzierten Websites ist auf den 31.12.2008 datiert.].

3 z. B. „EndNote“ – URL: <http://www.endnote.com>; „Citavi“ – URL: <http://www.citavi.com> (nur Windows)

4 z. B. „Refworks“ – URL: <http://www.refworks.com>; „CiteULike“ – URL: <http://de.citeulike.org>; „EndNote Web“ – URL: <http://www.endnoteweb.com>

tadaten frei im Internet zur Verfügung stehen, kann die Kollaboration in der Wissenschaft unterstützt werden – so lassen sich etwa Publikationslisten zu Projekten kooperativ pflegen und veröffentlichen. Darüber hinaus entstehen durch Software-Erweiterungen „Mischformen“ der beiden Prinzipien<sup>5</sup> sowie weitere Systeme mit speziellen Funktionen wie das Plug-in „Zotero“<sup>6</sup>, das im Webbrowser Mozilla Firefox integriert wird.

Weiterhin kann eine Unterscheidung getroffen werden zwischen Literaturverwaltungssystemen, die kommerziell angeboten werden (z. B. „Citavi“, „EndNote“, „Endnote Web“, „RefWorks“) und jenen, die frei und quelloffen im Internet zur kostenlosen Verfügung stehen (so genannte Open-Source-Anwendungen wie „Zotero“, „Connotea“, „CiteULike“, „Wikindx“<sup>7</sup> etc.).

## Funktionen von Literaturverwaltungssystemen

Literaturverwaltungssysteme lassen sich anhand ihrer Funktionen beschreiben. Dazu zählt die formale Erschließung von Literaturquellen in unterschiedlichen, systemabhängigen Erschließungsgraden anhand bibliografischer Daten wie Autor, Titel, Entstehungsort, Entstehungsjahr etc. Weiterhin bieten Literaturverwaltungssysteme die inhaltliche Erschließung von Quellen durch verschiedene Formen der Wissensorganisation wie beispielsweise die Vergabe von Schlagwörtern oder Kategorien, die Erstellung von Gruppen, Ordnern und themenspezifischen Bibliografien. Diese Form der Wissensorganisation kann system-abhängig um die Erschließung mit weiteren, eigenen Metadaten, Notizen, Anmerkungen erweitert werden – in diesem Sinne lassen sich z. B. ganze Zitatesammlungen verwalten. Zudem bieten einige der Anwendung weitere Zusatz-Funktionen, die nicht im Einzelnen aufgeführt werden können.<sup>8</sup>

Webbasierte Literaturverwaltungssysteme unterstützen teilweise das Social Tagging: Nutzer erschließen gemeinsam Internetquellen durch tags bzw. freie Schlagwörter, die ohne Regeln vergeben werden. Nach diesem Prinzip lassen sich durch andere Nutzer erschlossene Publikationen mit einer gezielten Suche nach tags finden bzw. können die Interessen anderer Nutzer „entdeckt“ werden.<sup>9</sup>

Literaturverwaltungssysteme ermöglichen entsprechend ihrer Erschließungsmethodik verschiedene Suchfunktionen und das Browsing in den gesammelten Referenzen und Notizen sowie die passende Erstellung von Bibliografien bzw. Literaturverzeichnissen. Die meisten Anwendungen verfügen über Schnittstellen zu Textverarbeitungsprogrammen (vorwiegend MS Word, Open Office und LaTeX) bzw. bieten Software-Ergänzungen mit dieser Funktion an (so genannte Add-ons). Die Zitierweise bzw. die Outputstyles, die sich in der Regel innerhalb verschiedener Fachdisziplinen unterscheiden (z. B. APA, MLA, Chicago), sind oftmals frei wählbar bzw. können um eigens erstellte Styles erweitert werden.

Unterstützt wird zudem die Ausgabe der bibliografischen Daten in verschiedenen Datenausgabeformaten, die einen Austausch bzw. den Import und Export von Metadaten zwischen verschiedenen Anwendungen ermöglichen. Literaturverwaltungssysteme unterstützen diese Vorgänge durch standardisierte Formate (BibTeX, RIS, Endnote, XML bzw. vor allem der von der Library of Congress entwickelte MODS-Standard etc.) und garantieren dem Nutzer dadurch die Möglichkeit des Wechsels zwischen verschiedenen Anwendungen oder Rechnern. Fachdatenbanken und Repositorien wie der Dokumenten- und Publikationsserver (Edoc) der Humboldt-Universität stellen die Metadaten zu den Publikationen in verschiedenen Formaten zur Verfügung und unterstützen damit die einfache Übernahme dieser Daten in Literaturverwaltungssysteme. Ähnliche Angebote gibt es auch von Seiten einiger Bibliotheken: So bietet etwa der OPAC der Universitätsbibliothek

Köln<sup>10</sup> den direkten Import von Metadaten zu Publikationen in das webbasierte Literaturverwaltungssystem „Bibsonomy“ an. Weiterhin können beispielsweise die Metadaten von Büchern auf den Webseiten des Internet-Versandhauses Amazon über ein Icon in der Adresszeile des Browsers direkt in das Literaturverwaltungssystem „Zotero“ übernommen werden oder über die ISBN-Nummer in die Anwendung „Citavi“ überführt werden.

## (Weiter-)Entwicklung von Literaturverwaltungsprogrammen an der HU

Auch an der Humboldt-Universität beschäftigt man sich mit der (Fort-)Entwicklung und Bereitstellung von webbasierter Literaturverwaltungssoftware. Zum Ersten wird am Institut für Bibliotheks- und Informationswissenschaft der HU derzeit an einem Update eines Bibliographiewerkzeugs für das an der Universität weit verbreitete Content-Managementsystem Plone gearbeitet.<sup>11</sup> Kultur- und Medienwissenschaftler verbinden in dem Multimedia-Projekt Liblia 2.0 „eine digitale Bibliothek und Literaturverwaltung mit der Funktionalität einer Online-Community“<sup>12</sup>. In diesen Entwicklungen wird gezielt den Bedürfnissen der Universitätsangehörigen Rechnung getragen.

In einer Zusammenfassung soll festgehalten werden, dass in diesem Überblick bei weitem nicht alle Funktionen dargestellt werden können. Literaturverwaltungssysteme gewinnen mit der Weiterentwicklung ihrer Funktionen bzw. Software-Erweiterungen und damit der vereinfachten Benutzbarkeit und der daraus entstehenden Mehrwerte im wissenschaftlichen Publikationsprozess zunehmend an Bedeutung. Perspektivisch lässt sich vermuten, dass die Grenze zwischen rein webbasierten und lokalen Anwendungen verschwimmen wird (so-

5 z. B. wird das System „EndNote“ durch „EndNote Web“ ergänzt

6 „Zotero“ – URL: <http://www.zotero.org>

7 „Wikindx“ – URL: <http://wikindx.sourceforge.net>

8 „Citavi“ bietet hier beispielsweise mit der „Aufgabenplanung“ und dem Management von Terminen eine zusätzliche Funktion und mit „Zotero“ können „Schnappschüsse“ von Webseiten gemacht werden, die sich editieren lassen.

9 z. B. „Connotea“ (URL: <http://www.connotea.org>) und „Bibsonomy“ (URL: <http://www.bibsonomy.org>)

10 „Kölner UniversitätsGesamtkatalog“ – URL: <http://kug.ub.uni-koeln.de>

11 vgl. Stöhr, Matti: Literaturverwaltung mit Plone, in: cms-journal, Nr. 31, S. 7–10

12 vgl. Liblia 2.0 – Eine digitale Bibliothek im World Wide Web – URL: <http://www.cms.hu-berlin.de/dl/multimedia/bereiche/mlz/projekte/projekt Datenbank/projekte/2007/liblia>

wie die Plattformunabhängigkeit der einzelnen Anwendungen erreicht werden sollte).<sup>13</sup> Der Rückgriff auf standardisierte Austauschformate für bibliografische Daten wird dabei im Vordergrund stehen, da verschiedene Literaturverwaltungssysteme mit den unterschiedlichsten Funk-

tionen parallel existieren werden und der unkomplizierte Datenaustausch für den Nutzer gewährleistet werden sollte.

Abschließend sei darauf hingewiesen, dass die Humboldt-Universität im Rahmen von Campuslizenzen die Literaturverwaltungssysteme „EndNote“ (Windows

und Mac)<sup>14</sup> und „Citavi“ (Windows)<sup>15</sup> anbietet sowie an Universitätsbibliothek und Computer- und Medienservice ein entsprechendes Informationsangebot bereitgehalten wird.<sup>16</sup>

<sup>13</sup> Beispielsweise wird für 2009 die Entwicklung einer Mac-Version von „Citavi“ sowie „Citavi Web“ angekündigt.

<sup>14</sup> siehe: <http://www.cms.hu-berlin.de/dl/software/lizenzen/endnote.html>

<sup>15</sup> siehe: <http://www.ub.hu-berlin.de/service/literaturverwaltung>

<sup>16</sup> siehe: [http://www.ub.hu-berlin.de/onlinekurs/ik\\_modul\\_d/02\\_01\\_literaturverwaltungsprog\\_01.html](http://www.ub.hu-berlin.de/onlinekurs/ik_modul_d/02_01_literaturverwaltungsprog_01.html)

## Personalia

Herr Matthias Schulz verließ den CMS am 15.2.2009, um eine Promotionsstelle an der Universität St. Gallen anzunehmen. Er arbeitete seit dem 1.5.2005 erfolgreich auf den Gebieten der Webserver- und Datenbankadministration in der Abteilung Systemsoftware und Kommunikation.

Am 31.3.2009 hat Heinz-Günter Kuper den CMS verlassen und wechselte zum Institut für Informatik. Er war seit 09.01.2006 in den Drittmittelprojekten Covcell und HyperImage beschäftigt und hat dort wesentliche Akzente in der Softwareentwicklung gesetzt und so zum guten Gelingen dieser Projekte beigetragen.

Seit dem 1.2.2008 ist Jana Kunze als Mitarbeiterin im EU-Projekt „dimekon“ eingestellt. Sie wird im Rahmen des Projektes, das vom MLZ gemeinsam mit der Abteilung für Personal und Personalentwicklung durchgeführt wird, Weiterbildungsmaßnahmen zur Vermittlung von digitaler Medienkompetenz entwickeln.

## Verabschiedung von Gisela Schulz

Am 1. Februar ist Gisela Schulz in den wohlverdienten Ruhestand gegangen. Sie begann ihre Arbeit an der Humboldt-Universität im Dezember 1980 im Videobereich des damaligen Zentrums für audiovisuelle Lehr- und Lernmittel. Damals wie heute bestimmten viele Videoaufzeichnungen aus allen Bereichen der Universität die Arbeit. Festveranstaltungen und andere Höhepunkte des Universitätslebens wurden dokumentiert. Ein weiterer Schwerpunkt dieser Jahre waren Unterrichtsaufzeichnungen von Schulstunden, in denen zukünftige Lehrer unterrichteten. Diese wurden entweder im Studio oder direkt in den Schulen aufgezeichnet. Weiterhin wurden für die damalige Sektion Fremdsprachen viele Lehrvideos, auch mit Spielszenen, zur Fremdsprachenausbildung für Englisch und Russisch aufgezeichnet.

1982 erfolgte dann der Umzug in das neugebaute Seminargebäude am Hegelplatz. Hier wurden viele Videoeinspielungen vom Regieraum direkt in die Seminarräume durchgeführt. Daneben

wurden Schnittplätze, zunächst analog, später digital, für studentische Projekte zur Verfügung gestellt. Die ersten digitalen Schnittcomputer an diesen Arbeitsplätzen waren 1996 Casablanca-Schnittsysteme mit 2 GB Speicherplatz. Weiterhin spielten Aufnahmen von Lehrvideos eine wichtige Rolle.

1998 wurde dann die Zentraleinrichtung für audiovisuelle Lehrmittel, wie sie mittlerweile hieß, aufgelöst und die Aufgaben sowie die Mitarbeiterinnen und Mitarbeiter wurden weitgehend dem Rechenzentrum zugeordnet. So kam Gisela Schulz in die neugegründete Multimediadienstleistungs-Abteilung, blieb aber an ihrem vertrauten Arbeitsort im Seminargebäude. Dieser veränderte sich erst mit dem Umzug des, mittlerweile in Computer- und Medienservice umbenannten, Rechenzentrums nach Adlershof. Im Erwin Schrödinger-Zentrum war ein moderner Videobereich mit Aufnahme- und Tonstudio entstanden. Die rasante Weiterentwicklung der Technik ließ die Arbeit immer interessant und

spannend bleiben und schaffte neue Herausforderungen. Außenaufnahmen gewährten Einblicke in Bereiche der Universität, die sonst der Öffentlichkeit nicht zugänglich sind. Gisela Schulz selbst bezeichnet ihre Arbeit als „vielseitige, interessante Arbeit, die fast gleich geblieben ist, allerdings mit sehr viel komfortableren Arbeitsmitteln“.

Gisela Schulz ist für ihre Einsatzbereitschaft bekannt. Aufnahmen auch außerhalb der normalen Arbeitszeiten waren für sie nie ein Problem. Auch unter den Kolleginnen und Kollegen hatte sie immer ein offenes Ohr für Fragen und Probleme und war stets hilfsbereit.

Für ihre umfangreichen und vielfältigen Leistungen bedanken wir uns auf das Herzlichste. Wir wünschen ihr vor allem Gesundheit und Zeit für die Dinge, die während des Berufslebens zu kurz gekommen sind. Wir hoffen, dass die Verbindung zu Frau Schulz bei möglichst vielen fachlichen und festlichen Gelegenheiten bestehen bleibt.



## Verabschiedung von Herrn Kirchner

Am 28.04.2009 hat sich Herr Uwe Kirchner von den Kolleginnen und Kollegen des CMS mit einer kleinen Feier verabschiedet. Er hat die Humboldt-Universität nach fast 33 Jahren verlassen und wird nun den wohlverdienten Ruhestand genießen.

Nach Abschluss seines Studiums an der TH Magdeburg als Diplom-Mathematiker 1970 arbeitete er zunächst im Rechenzentrum des Metallurgie-Außenhandels Berlin in der Programmentwicklung. Am 12.12.1976 begann Uwe Kirchner seine Tätigkeit an der Humboldt-Universität. Er hat viele Namensänderungen der Einrichtung mitgemacht – Organisations- und Rechenzentrum (ORZ), Rechenzentrum (RZ), Computer- und Medienservice (CMS). In der Abteilung I, heute Datenverarbeitung in der Verwaltung, war er von Anfang an, und auch hier gab es unterschiedlichste Bezeichnungen. Eingestellt wurde er als Problemanalytiker, verlassen hat er uns als Administrator für das Windowsnetzwerk der Verwaltung. In den ersten Jahren an der HU war er u. a. an der Entwicklung und Programmierung eines Studierenden- und Stipendienabrechnungssystems beteiligt. Mit dem Einsatz der ersten Personalcomputer an der HU (EC 1834,

AC 5120 und PC 1715) programmierte er das „Formblatt 1514“ eines Forschungsabrechnungssystems – damals ein Dauerthema auf vielen Dienstberatungen. Als Folge der Wirtschafts-, Währungs- und Sozialunion 1990 wurde die Gehaltsabrechnung der HU auf ein neues Abrechnungssystem des Berliner Senates umgestellt. An dieser Umstellung war Herr Kirchner maßgeblich beteiligt. Auch die Übernahme der Personaldaten in ein neues DV-System (HISSVA) gehörte zu seinen Projekten. Er entwickelte mehrere Programme für die Gehaltsstelle der HU, von denen auch heute noch einige genutzt werden. Nach 1995 änderte sich das Aufgabengebiet von Uwe Kirchner – er musste allmählich Abschied nehmen von der reinen Programmierung und sich in die Administration des Netzwerkbetriebssystems Banyan Vines und die PC-Installation einarbeiten. Seit 2004 arbeitete er in der Arbeitsgruppe Windowsnetzwerke in der Verwaltung – er war der Spezialist für die Virenbekämpfung, betreute die Hotline und griff schon gern mal selbst zum Schrauber... Er war unser Sammler von Programmen, von Software und PC-Ersatzteilen. Wenn man ihn um Hilfe oder einen Gefallen bat, war er sofort zur Stelle.

Die Kolleginnen und Kollegen der Verwaltung werden ihn als hilfsbereiten CMS-Mitarbeiter vermissen.

„Jedem Anfang wohnt ein Zauber inne“ [Hermann Hesse] – wir wünschen Uwe Kirchner viel Glück für den nächsten Lebensabschnitt. Wir kennen ihn gut und wissen, dass er viele Interessen hat – Langeweile wird er wohl nicht haben.